

CORRECTED  
VERSION\*

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : <b>C12N 5/00, 15/13, G01N 33/53</b>	A1	(11) International Publication Number: <b>WO 99/47643</b> (43) International Publication Date: 23 September 1999 (23.09.99)
(21) International Application Number: PCT/US99/05944		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
(22) International Filing Date: 18 March 1999 (18.03.99)		
(30) Priority Data: 60/078,432 18 March 1998 (18.03.98) US		
(71) Applicant (for all designated States except US): QBI ENTERPRISES, LTD. [IL/IL]; Weizmann Scientific Park, P.O. Box 741, 74106 Nes-Ziona (IL).		
(72) Inventors; and		
(75) Inventors/Applicants (for US only): GUDKOV, Andrei [US/US]; 700 Sunset Lane, Glencoe, IL 60022 (US). KONDRATOV, Roman [RU/RU]; 901 South Ashland Avenue, Chicago, IL 60607 (US).		Published <i>With international search report.</i>
(74) Agents: KOHN, Kenneth, I. et al.; Kohn & Associates, Suite 410, 30500 Northwestern Highway, Farmington Hills, MI 48334 (US).		

(54) Title: SELECTION SUBTRACTION APPROACH TO GENE IDENTIFICATION

(57) Abstract

The present invention provides a method of creating expression libraries by marking desired clones with a unique tag. An expression library including individual clones marked with a unique tag sequence is also provided. Also provided is a method of isolating genetic elements associated with complete or partial cell growth suppression or cell death by isolating clones from an expression library which are lost in the cell population. A method of selection subtraction is provided by tagging a clone in an expression library, cloning the tagged clone into a vector, delivering the tagged clone to a target cell, and comparing tags before and after selection whereby toxic genes and the attached tags disappear.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

**SELECTION SUBTRACTION APPROACH TO GENE IDENTIFICATION****CROSS REFERENCE TO RELATED APPLICATIONS**

This application is a conversion of a United States Provisional Application filed March 18, 1998, Serial No. 60/078,432.

**BACKGROUND OF THE INVENTION****10 TECHNICAL FIELD**

The present invention relates to the creation of "tagged" expression libraries, wherein each cDNA clone is marked by a unique TAG. The present invention further relates to the identification of genes using different methods of TAG-based selection-subtraction that identify genes associated with complete or partial growth suppression, with cell death or with any other selectable phenotype, by recognition and comparison of their tagged clones before and after selection. Furthermore, the present invention relates to the elaboration of novel subtraction techniques and a novel type of expression vector, a retrophage.

**25 BACKGROUND ART**

There is no established and reliable technology available for functional isolation of growth suppressive genetic elements. Method "SETGAP" (Pestov et al, 1994) is based on expression libraries but has the limitation of only identifying subsets of growth inhibitory clones that are strong enough to block DNA synthesis after induction and are not too toxic to kill the cells after expression induction. It would be useful to have a method that does not have these limitations.

Expression selection is a widely used method for identification and cloning of genes which involves the use of expression gene libraries. However, there is a problem with existing expression libraries in that they are 5 significantly redundant. It would be useful to obtain expression libraries from complex libraries, in which each unique clone is tagged to facilitate isolation.

Functional identification of cytotoxic and cytostatic genetic elements is also an important and 10 challenging task. Establishment of a reliable method allowing expression selection of such elements could greatly facilitate discovery of new tumor suppressor genes that are mutated, deleted or inactivated by other means in processes of tumor development and progression. On the 15 other hand, cytotoxic genetic elements with certain cell specificity (for example, directed against cancer cells), may serve as efficient tools in gene therapy applications. Although there are powerful methods to identify positive regulators of cell growth, the approaches to isolation of 20 toxic genes are limited to the laborious methods of genetic analysis based on karyotypic alterations in cancer cells and GSE and anti-sense RNA technologies.

The methods used with and the utility of the present invention can be shown by the following non-limiting examples and accompanying figures. 25

#### SUMMARY OF THE INVENTION

In accordance with the present invention, there 30 is provided a method of creating expression libraries by marking desired clones with a unique tag. An expression library including individual clones marked with a unique tag sequence is also provided. Also provided is a method of isolating genetic elements associated with complete or 35 partial cell growth suppression or cell death by

isolating clones from an expression library which are lost in the cell population. A method of selection subtraction is provided by tagging a clone in an expression library, cloning the tagged clone into a vector, delivering the tagged clone to a target cell, and comparing tags before and after selection whereby toxic genes and the attached tags disappear.

**10****DESCRIPTION OF THE DRAWINGS**

The advantages of the present invention will be readily appreciated as the same becomes better understood by reference to the following detailed description, when considered in connection with the accompanying drawings,  
**15** wherein:

**20** Figure 1 shows the general principle of selection-subtraction approach (SSA); during passaging, "toxic" cDNA clones disappear from the recipient cells transduced with cDNA expression library and are subsequently identified by subtraction;

**25** Figure 2 shows the general principle of selection-subtraction approach based on the subtraction of individual TAGs that mark each library clone;

**30** Figure 3 shows the structure of a novel type of expression vector combining the features of  $\lambda$  phage and retroviral vector, a retrophage;

**35** Figure 4 shows the structure of adaptors used for the tagged library construction; a & b are adaptors for the tagged p53 fragmented cDNA library; c is an adaptor for tagged full-length cDNA library; d is a TAG-containing primer for full-length cDNA library;

Figure 5 shows the method of addition of TAGs to cDNA clones by PCR;

5 Figure 6 shows the method of addition of TAGs to cDNA clones during reverse transcription;

Figure 7a shows the structure of retroviral vector with incorporated T7 promoter sequences;

10 Figure 7b demonstrates the implication of this vector in a design of full-size tagged library construction; a is the structure of tag; b is the structure of a 5' adaptor; c is the structure of a 3' oligonucleotide used for first strand synthesis; d are the  
15 primers for TAG rescue/amplification;

Figure 8 presents the method for identification of differences in full-length library composition before and after selection, using hybridization to cDNA microarrays;

20 Figure 9 shows the TAG-based subtraction method using Mung Bean Nuclease treatment;

25 Figure 10 shows the TAG-based subtraction method using SSH;

Figure 11 describes the RNase H mediated subtraction method;

30 Figure 12 presents the scheme of construction of a randomly fragmented tagged p53 single gene library;

35 Figure 13 presents the general strategy of a tagged cDNA library (either single gene or total) construction and screening;

Figure 14 demonstrates the disappearance of "toxic" full-length p53 cDNA clone from transduced 041 and 10(1) cells during passaging, as detected by Southern blot analysis with the corresponding tag sequences;

5

Figure 15 presents the method of identification of "toxic" cDNA clones by arraying the library clones on nylon membrane and subsequent differential hybridization of these membranes to the pools of TAGs rescued from cells  
10 before and after selection (passaging);

Figure 16 presents the general strategy of tagged cDNA library (either single gene or total) construction and screening using the RNase H-based approach;

15

Figure 17 demonstrates the results of RT-PCR from RNA before and after RNase H-based tag subtraction procedure; bottom panel shows isolation of the inserts marked by differentially represented tags by PCR using  
20 vector sequences flanking the insert as primers;

Figure 18 illustrates technical steps of SSH-based procedure of tag subtraction; top panel shows polyacrylamide gel fractionation of the isolated tags  
25 treated with the indicated restriction enzymes; bottom panel represents an agarose gel with the products obtained after 40 cycles of PCR on the indicated templates;

Figure 19 demonstrates the efficacy and  
30 specificity of recovery of the library-derived inserts by RT-PCR using vector-specific primers;

Figure 20 presents PCR products of recovered inserts at different stages of selection of three  
35 retroviral cDNA libraries in three different cell types;

Figure 21 illustrates elimination of growth suppressive clone encoding p21/WAF1 sequence during passaging of target library-transduced cells monitored by Southern hybridization; and

5

Figure 22 presents Southern blot hybridization with the results of RNase H-based subtraction procedure. Left panel shows elimination of 2A-specific tag during cell selection, while the right panel presents strong 10 enrichment of this tag in RT-PCR product synthesized on RNA after subtraction with the tags isolated after selection.

15

#### DETAILED DESCRIPTION OF THE INVENTION

The present invention provides an alternative method of identification of gene sequences whose expression in cells elicits growth suppression, apoptosis or any other selectable phenotype. It is based on the 20 isolation of those particular clones from an expression library that are specifically lost in cell population transduced with the expression library under certain conditions of selection, i.e., simple passaging or sensitization to certain cytotoxic stimuli, by using gene 25 subtraction procedures.

Standard molecular biology techniques known in the art and not specifically described are generally followed as in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Springs Harbor Laboratory, New York (1989, 1992), and in Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley and Sons, Baltimore, Maryland (1989). Additionally, standard methods in immunology known in the art and not specifically described are generally followed as in

Stites et al. (eds), *Basic and Clinical Immunology* (8<sup>th</sup> Edition), Appleton & Lange, Norwalk, CT (1994) and Mishell and Shiigi (eds), *Selected Methods in Cellular Immunology*, W.H. Freeman and Co., New York (1980).

5       Delivery systems (vectors) for delivering the expression cassette to the target cell are known in the art and a nonexhaustive list is provided herein below. Of particular interest are viral vectors as discussed below. Alternatively, a cell based system is used as a method of  
10 site-specific delivery (targeting) to the tumorogenic cells. The cells are generally selected from cells that shed and/or release the expression cassette to be incorporated by the tumorogenic cells as is known in the art. These cell are selected from cells that will "home"  
15 to the site of the tumorogenic cells. Alternatively, other methods know in the art for site specific delivery of cells are used to deliver the cells *in situ*.

Vectors which comprise the DNA encoding for a clone of the expression library are also provided by the  
20 present invention. The vectors can be constructed by those skilled in the art and should contain all expression elements necessary to achieve the desired transcription of the sequences. Other beneficial characteristics can also be contained within the vectors  
25 such as mechanisms for recovery of the nucleic acids in a different form. Phagemids are a specific example of such beneficial vectors because they can be used either as plasmids or as bacteriophage vectors. Examples of other vectors include viruses such as bacteriophages,  
30 baculoviruses and retroviruses, DNA viruses, cosmids, plasmids, liposomes and other recombination vectors. The vectors can also contain elements for use in either procaryotic or eucaryotic host systems. One of ordinary skill in he art will know which host systems are  
35 compatible with a particular vector.

The essential materials and experimental steps of the present approach are described below.

1. Expression Libraries

5 The approach is applicable to a wide variety of expression libraries including plasmid- and virus-derived expression systems in eukaryotes and prokaryotes. The source of genetic material may involve mixtures of cDNAs cloned in sense or anti-sense orientation, mixtures of  
10 fragments of genomic DNA or cDNA, fragments of individual cDNAs and mixtures of artificial sequences.

2. Library Delivery

15 Expression libraries are delivered to the test cells by an appropriate gene transfer techniques, including various methods of DNA transfection, infection by vector viruses or bacteriophages.

3. Selection Conditions

20 Growth suppressive clones will be gradually lost during propagation of the cell population carrying the library. Such clones may decrease growth rate or kill the cells under normal conditions of growth or they may be growth suppressive under specific conditions, including  
25 sensitization of cells to certain physical, chemical or biological treatments that have no or low effect on the majority of cells in library-carrying cell population. See Figure 1.

30 4. Screening Procedures

Expression library or the library inserts are rescued from the cells before and after selection. The rescue procedure is determined by the properties of the library and may be done by PCR on total cellular DNA, RT-  
35 PCR on cellular RNA, direct transformation of competent cells (episomal libraries) or virus rescue (libraries in

retroviral, adenoviral or other viral vectors). The proportion of growth-suppressive sequences is expected to be decreased in the rescued libraries or library inserts after the selection. These sequences are isolated by using  
5 one of the gene subtraction techniques. See Figure 1.

Specifically, library-derived sequences of the resulting (rescued) mixture of molecules is subtracted from the original mixture of sequences either derived from the original library or rescued from the cells before  
10 selection. The complexity of the subtracted material is estimated. If the number of isolated clones is too large for individual testing, the material is subjected to a second round of selection. This procedure is repeated until the number of clones becomes low enough to allow  
15 individual testing. See Figure 1.

Technically, subtraction may be done in different ways, including manipulations with the sequences of the genes themselves and using a newly developed tagged library screening technique. During construction of cDNA  
20 libraries, each clone is marked with a short 80 bp random DNA fragment (TAG).

A tagged library is cloned in retroviral vector and delivered to target cells. Total cell RNA is isolated immediately after introduction into the cell and after  
25 application of selection conditions, as aforementioned. During selection, cells expressing "toxic" genes are eliminated from the population and TAGs of these "toxic" genes will disappear. Comparison is made of TAGs before and after selection, (for example, using PCR-based  
30 subtraction methods), and disappearing TAGs marking putative cytotoxic genes are isolated. These TAGs are then used as probes to identify the cytotoxic and cytostatic genetic elements which they mark in the library. (See Figure 2).

The use of TAGs has a number of advantages: (1) the TAGs have uniform size and are designed in such a way as to facilitate their isolation by PCR or other methods; (2) each clone in the library is marked by a unique TAG, 5 with the full-length, truncated or anti-sense variants of the same gene marked with different TAGs. Subtraction allows identification of those particular clones that are biologically functional; and (3) this method can be applied for the detection of "toxic" sequences regardless 10 of their origin since all subtraction procedures are done in standard conditions (full length cDNA, GSE, etc.).

##### 5. Isolation of Growth-suppressive Genetic Elements with Cell-Specific Activity

15 To isolate genetic elements that are growth-suppressive for certain cell types (tumor cells, virus-infected cells, cells of certain tissue, etc.) that have no or low effect on other cell types, the selection procedure is applied in parallel to the target cells and 20 to the cells of other types that are not supposed to be affected by the isolated elements (i.e., transformed cells and non-transformed cells or similar origin, etc.). The resulting mixtures of molecules obtained by selection subtraction approach (SSA) procedures are compared, to 25 isolate those sequences that are specifically lost during selection of the target but not control cells. Specifically, such sequences are isolated by subtraction of the material obtained from control cells from the material obtained from the target cells. Different 30 subtraction procedures can be applied for the isolation of the desired tags, including direct hybridization comparison of the populations of pre- and post-selected tags, RDA, SSH and any other methods allowing for the comparison of two populations of nucleic acid sequences.

6. Expected Growth Suppressive Genetic Elements and their Possible Applications

This method has two general applications: (I) isolation of growth suppressive molecules (DNA, RNA or polypeptides) with specificity against certain cell types; and (II) identification of new genes involved in negative growth regulation (i.e., tumor suppressive genes, genes involved in control of apoptosis, cellular senescence, sensitivity to chemical, physical or biological treatments, etc.).

7. Other Types of Selections of Biologically Active Genetic Elements

It is noteworthy that the range of applications of SSA is not limited to the isolation of cytotoxic or cytostatic elements. It can be used for isolation of genetic elements that induce any cell phenotype that can be used as a selective trait to enrich or exhaust cell population (i.e. expression of cell surface antigens, alterations in cell adhesion, cell size, etc.).

#### GENERAL METHODS

General methods in molecular biology: Standard molecular biology techniques known in the art and not specifically described are generally followed as in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Springs Harbor Laboratory, New York (1989, 1992), and in Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley and Sons, Baltimore, Maryland (1989). Polymerase chain reaction (PCR) is carried out generally as in *PCR Protocols: A Guide to Methods and Applications*, Academic Press, San Diego, California (1990). Reactions and manipulations involving other nucleic acid techniques, unless stated otherwise, are performed as generally

described in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Springs Harbor Laboratory, New York (1989) and methodology as set forth in United States patents 4,666,828; 4, 683,202; 4,801,531; 5,192,659 and 5 5,272,057 and incorporated herein by reference.

#### SPECIFIC METHODS

Expression libraries that can be used for SSA

10 The developed approach is applicable to a wide variety of expression libraries including plasmid- and virus-derived expression systems in eukaryotes and prokaryotes. The source of genetic material may involve mixtures of cDNAs cloned in sense or anti-sense  
15 orientation, mixtures of fragments of genomic DNA or cDNA, fragments of individual cDNAs and mixtures of artificial sequences. However, the use of conventional expression libraries for SSA could be complicated by redundancy of the libraries that might carry similar  
20 sequences in both functional and non-functional forms and by difficulties of rescuing and subtraction of the inserts of variable length. These obstacles could be overcome by the use of Tagged libraries.  
25 New "retrophage" expression vector

The phage lambda-based vectors have significant advantages over plasmid vectors in preparation and screening of complex libraries. The efficacy of ligation of the inserts into a phage vector depends significantly  
30 less on the insert size (within the regular range of lengths of the majority of cDNAs). Phages can be plated at much higher density than bacterial colonies thus increasing the efficiency of screening of large libraries. Hybridization screening of the prints of phage  
35 plaques on nylon or nitrocellulose is a far more

sensitive and reliable technique than that of the bacterial colonies. Prints of plaques of phage libraries are highly reproducible and can be considered cDNA arrays allowing one to simultaneously analyze expression of 5 numerous genes in a single hybridization.

At the same time, retroviral expression vectors have a series of advantages specifically valuable for generation and screening of expression libraries (cDNA, GSE, antisense RNA, etc.). They represent the most 10 efficient delivery vehicle that brings the gene of choice into the cell and determines its integration and stable expression in the genome at different levels within the range of expression of cellular genes.

The advantages of phage and retroviral vectors 15 were combined by creating a new retrovirus expression vector inside the phage DNA. The schematic presentation of this new vector is shown at the Figure 3. The new vector consists of the λgt11 phage with a LXSN retrovirus vector inside. The polylinker inside LXSN was substituted for two 20 SfiI sites with different cohesive ends to make it possible for oriented cloning of the inserts and to avoid self-ligation of the two arms of the phage DNA. λgt11 can accept fragments of DNA up to 7.2 kb in length. The size of provirus part is about 3 kb. Thus, the cloning capacity 25 of the new vector is up to 4 kb.

To prove the efficacy of the created vector, its ability to produce infectious retroviruses after transfection into packaging cells was directly compared with that of the LXSN plasmid vector and found it to be 30 only 2 to 3 fold lower if calculated per microgram of DNA. The phage/virus vector can be very useful for preparation of large expressing cDNA libraries with millions of recombinant clones. To increase the cloning capacity of the phage/virus vector we made a new 35 modification was made to the original vector, named

lgtLXL, which lacks selectable marker gene and has cloning limit up to 5 kb.

## 5 METHODS OF PREPARATION OF TAGGED LIBRARIES

### Nature of tags

As tags one can use any highly divergent sequences, the complexity of which significantly exceeds the complexity of the library. For example, random **10** oligonucleotides or fragments of low redundant genomes (i.e. DNA of prokaryotes) of a reasonable size range (10-1000 bp) can be used as tags. The choice of tags is dictated by the method of tag isolation, subtraction and identification (see below). For example, random **15** oligonucleotides 80 bp long flanked by short (approximately 20-30 bp) sequences carrying the appropriate restriction sites and suitable for efficient PCR amplification (Figure 4) were used as tags. Tags are introduced in the library-carrying vectors at the sites **20** at which they cannot interfere with the expression or delivery.

### Introduction of Tags into expression libraries

Oligonucleotides can be introduced into the **25** library at different stages of library preparation including the following options:

- 1) Tags are introduced into the population of vector molecules prior to library preparation, and a highly diverged population of individually tagged vectors is **30** used for cloning of inserts;
- 2) Tags are introduced in already prepared libraries by cloning the Tagging DNA fragments into the defined site of the existing libraries;
- 3) Tags are added to the inserts prior to cloning into **35** expression vectors.

Among the three above options, the third one has clear advantages since it ensures that every individually generated clone is marked by a specific tag sequence. Two examples of Tagged library construction are shown below.

5 In one of them, oligonucleotide tags are added to cDNA fragments by ligation followed by PCR amplification (Figure 5). Another example shows 80-nucleotide long Tags incorporation during the first stage of cDNA synthesis. In this method, Tag sequences are included as part of the  
10 primers used for reverse transcription (Figure 6).

#### Library delivery and selection

The choice of methods for library delivery and selection depends on the nature of the expression vector,  
15 cell type and experimental needs. Plasmid libraries can be delivered to target cells by transfection, lipofection, electroporation and other available techniques. Viral vectors can be first delivered to packaging cells to convert the library from the plasmid  
20 form into the mixture of viral particles, which then is used for the infection of target cells. Examples of construction and delivery of retroviral libraries are shown below; methods of their delivery to target cells were described earlier (Gudkov & Roninson, 1996).

25 Selection of putative cytotoxic and cytostatic clones may involve simple propagation of library-transduced cell populations, thus allowing the cells to pass through several divisions to reduce the proportion of growth arrested or dead cells in the final population.  
30 Isolation of the clones that are not cytotoxic or cytostatic by themselves, but which display their growth suppressive activity under certain conditions of cell maintenance or treatment requires application of the appropriate selection. Such selection should be  
35 noncytotoxic or low-cytotoxic by itself and should not lead to a significant cell death or growth arrest that

might jeopardize identification of the desired library clones cooperating with the applied treatment in facilitating cell death or growth inhibition. For example, genetic elements restoring the activity of the p53 pathway by blocking p53 inhibitors (i.e., papilloma virus protein E6 or cellular p53 inhibitor MDM2, see examples below) are expected to cause cell sensitization to DNA-damaging treatments since p53 plays a role of negative growth regulator that is activated under conditions of genotoxic stress.

It is noteworthy that the range of applications of SSA is not limited to the isolation of cytotoxic or cytostatic elements. In principle, it can be used for isolation of genetic elements that induce any cell phenotype that can be used as a selective trait to enrich or exhaust cell population (i.e. cell surface antigens, cell adhesion, cell size, etc.).

**20 METHODS OF IDENTIFICATION AND CLONING OF GROWTH SUPPRESSIVE CLONES**

A. Methods based on subtraction of the library inserts

SSA can be applied to regular untagged libraries. 25 In this case, the inserts are isolated shortly after library delivery (before selection) and after selection and the populations of inserts are compared with each other to identify the inserts that were eliminated during selection.

30

Recovery of the inserts

Inserts can be recovered from the cells by PCR, using sequences flanking the tags as primers. Either cellular DNA or RNA can be used as templates. In the 35 latter case, cDNA should be synthesized first using oligo(dT), random oligonucleotides or specific

oligonucleotides corresponding to vector sequences as primers. For the libraries constructed in episomal vectors or in the vectors allowing to excise the inserts by Cre Lox recombinase (Russ et al., 1996) transformation 5 of competent bacterial cells by isolated plasmid DNA could be used instead of PCR. Example of the recovery of inserts from the cells transduced with retroviral libraries can be found in earlier published papers (Gudkov et al., 1994; Gudkov & Roninson, 1996). Rescuing 10 of the inserts may involve the stage of amplification of the isolated sequences by RNA-polymerase reaction. To make it possible, the expression vectors used for the library preparation can be supplied with the appropriate sequences serving as promoters of efficient prokaryotic 15 RNA polymerases (i.e., RNA polymerase 7, see Figures 7 and 8). RNA is synthesized using PCR products with T7 promoter as templates.

Comparison of pools of inserts using hybridization with 20 DNA microarrays

Comparison of the contents of two populations of inserts isolated from the cells before and after selection can be done using differential hybridization with microarrays containing, for example, a large number 25 of cDNAs representing individual genes or the inserts from the library used for the screening. Hybridization probes are prepared from the two pools of DNA inserts for comparison (i.e., by labeling with two different fluorescent nucleotide precursors; labeling can be done 30 in different ways, including random primed DNA-polymerase reaction or T7 RNA polymerase reaction) that are then used for hybridization with an appropriate DNA chip (Schena et al., 1995). Individual DNA probes showing differential hybridization are picked and used for 35 further characterization.

Isolation of the inserts eliminated during selection by the SSH technique

The comparison of the two pools of DNA inserts and isolation of differentially represented sequences can be  
5 done using a variety of gene subtraction protocols including previously described methods of RDA (representation differential analysis, Lysitsin et al., 1993) or SSH (suppression subtraction hybridization, Diatchenko et al., 1996). Both PCR-based methods require  
10 restriction digestion of the isolated inserts by frequently cutting enzymes, ligation with synthetic adaptors and PCR amplification. The resulting fragments of differentially represented inserts are then used as probes to clone their corresponding full size inserts.

15

Identification of the differentially represented inserts by colony or plaque hybridization

An expression library in the form of a mixture of recombinant plasmids or phages is delivered to  
20 host bacterial cells. The prints of the resulting colonies or plaques on nitrocellulose or nylon are then prepared according to standard protocols and hybridized with the recovered inserts. Comparison of hybridization results obtained with pre- and post-selection inserts  
25 allows one to pick differentially represented clones for further characterization.

#### B. Methods based on subtraction of Tags

A set of different gene subtraction procedures can  
30 be applied to identify and isolate differentially represented tags. These procedures involve both previously designed techniques, which have been adapted to the specific needs of tag subtraction, and novel approaches. Some of them are described below.

35

Tag subtraction, using modified RDA technique and single stranded DNA-specific nuclease

The principles of this approach are shown in Figure 9. For this method, tag sequences should be flanked by cleavage sites for restriction enzyme(s) (i.e., HindIII and EcoRI, as shown in Figures 4 and 8). Flanking sequences are removed from tags isolated from the cells after selection (TAG B) by digestion with the appropriate restriction enzymes and the fragments representing tag sequences are gel-purified. TAGs from the cells before selection (TAGs A) are also isolated and used uncut. An excess of TAGs B DNA is mixed with TAGs A, precipitated, dissolved in hybridization buffer and hybridized under standard conditions to reach saturated hybridization. The resulting mixture is diluted and treated with a nuclease specific for single-stranded DNA (i.e., Mung Bean nuclease) to remove non-annealed flanking sequences. DNA molecules that retained flanking adaptor sequences are then amplified by PCR. This procedure is repeated several times and the resulting material is used as a probe to isolate clones from the original library containing differentially represented inserts.

**25 Tag subtraction using modified SSH technique**

For this method, sequences that flank tags from the left and right sides should be different and should contain cleavage sites for two different restriction enzymes (i.e., HindIII and EcoRI, as shown in Figure 10). Driver DNA (TAGs B isolated from cells after selection) is cut with both enzymes and carefully gel-purified. Tester DNA (TAGs A, tags from the library prior to selection) is divided in two portions, which are cut with each of the restriction enzymes separately and also purified from undercut material. After that, a small amount of tester TAGs, of each type, are mixed with an

excess of driver TAGs in two separate tubes, they are precipitated, dissolved in hybridization buffer, and then annealed at standard conditions. Then the hybridization mixtures from two separate tubes are combined together 5 and further annealed. The resulting material is used for PCR. Only those sequences that are lost or underrepresented in the post-selection population of tags are preferentially amplified (Figure 18).

**10 RNAseH-based subtraction method**

A new subtraction procedure was developed specifically for SSA technique that is based on the use of RNase H for isolation of differentially represented sequences. The general principle of the method is 15 illustrated by Figures 11 and 16. In this method, only tags from the post-selected library (TAGs B) are isolated by PCR. These tags are used as a tool to subtract from the original RNA population those sequences that have not been lost during selection. The mixture of TAGs B 20 oligonucleotides isolated by PCR is digested by the appropriate restriction enzymes to remove uniform flanking adaptor-derived sequences and tag-related fragments are gel-purified. The resulting DNA is denatured, mixed with RNA isolated from the cells prior 25 to selection, dissolved in hybridization buffer and annealed under standard hybridization conditions of high stringency. The resulting mixture of RNA, DNA and RNA:DNA hybrids is then dissolved in an appropriate buffer and treated with RNase that destroys RNA in RNA/DNA hybrids. 30 Reverse transcription is then done on the remaining RNA using vector-specific antisense primers to specifically copy library-derived transcripts. The remaining tags are then amplified by PCR using flanking vector or adaptor sequences as primers. Isolated tags are used as 35 hybridization probes for detection of their corresponding clones from the original library. This subtraction

procedure allows one to directly synthesize a full size inserts by using as sense primers vector sequences located upstream of the insert. In this case, the resulting material is directly cloned into expression 5 vector and tested for biological activity (Figure 17).

It should be noted that a similar procedure can be applied to isolate differentially represented inserts from regular untagged expression libraries.

**10** Identification of differentially represented tags by colony or plaque hybridization

An expression library in the form of a mixture of recombinant plasmids or phages is delivered to host bacterial cells. The prints of the resulting colonies or 15 plaques on nitrocellulose or nylon are then prepared according to standard protocols and hybridized with the recovered tags. Comparison of hybridization results obtained with pre- and post-selection tags allows one to pick differentially represented tags for further use as 20 probes to isolate their corresponding library clones by conventional hybridization screening procedures.

Isolation of the inserts based on the results of tag subtraction

25 There are two main possibilities for isolating library clones containing identified tags. One involves the use of tags as hybridization probes to screen the original library by conventional colony- or plaque-hybridization. The other approach is based on PCR, using 30 tag-derived antisense primer and vector-derived sense primer corresponding to the sequence upstream of the inserts.

#### Secondary screening procedures

A single round of SSA may not necessarily result in isolation of a pure set of tags or inserts that are differentially represented in two pools of DNA (prior and after selection); the mixture obtained after the first round may be enriched with the desired sequences but could also be significantly contaminated with the sequences equally represented in both DNA samples. To reach better purification, the material obtained after the first round will be used to isolate the corresponding clones from the original library and the resulting sublibrary will be subjected to a second round of selection instead of carrying out individual characterization of the isolated inserts after the first round.

Isolation of clones that are specifically toxic for certain cell type may require additional selection of the isolated mixture of clones in the cell type of different origin in order to define those clones that have selective biological activity against the cells of choice.

#### Testing of biological activity of recombinant clones isolated by SSA

Recombinant clones isolated by SSA will be individually tested for their biological activity. They will be delivered by transfection or infection into target cells and the proliferation of transduced cells will be monitored under conditions used for the selection. Vectors carrying biologically inactive inserts or already known cytotoxic or cytostatic genetic elements (i.e. 2A gene of poliovirus or p21/waf1 gene that block cap-dependent translation and induce G1 arrest, respectively) will be used as controls for comparison. Relative efficiency of delivery of different constructs will be controlled (i.e. by co-transfection with the

LacZ- or GFP-expressing constructs). The behavior of the transduced cells will be monitored using either GFP fluorescence or selection in the presence of the appropriate selecting agent (G418, hygromycin, puromycin, 5 phleomycin, etc., depending on the vector). Clones with the confirmed biological activity will be picked for further detailed analysis.

#### EXAMPLES

10 **Example 1.** Construction of tagged retroviral libraries expressing fragments of individual cDNAs

cDNA inserts were gel-purified and partially digested by DNase I in the presence of Mn<sup>++</sup> (1μg DNA with 0.01 U of DNase I in 10mM MnCl<sub>2</sub>, 50mM Tris-HCl at 16°C at 15 different time points (between 1 and 5 min)). The ends were filled in by a mixture of Klenow fragment and T4 DNA polymerase, and the cDNA fragments ranging in size between 200 bp and full-length cDNA were ligated with two synthetic adaptors, one encoding FLAG peptide in-frame 20 with the ATG codon in Kozak's consensus, and containing EcoRI site and the other containing three stop codons in all reading frames followed by random 80 nucleotides (tag), flanked by EcoRV sites inside short "arms" for future PCR amplification, and BamHI site for cloning (see 25 Figure 4). The ligated material was amplified by PCR (10 cycles) with the following primers: sense - ATGAATTCTACCGCGGTGGCATGG (specific for 5' adapter) and antisense - ATGGATCCAGTTCCCTCGGTC (specific for 3' adapter). PCR products were fractionated and gel-purified 30 from 2% agarose, cut with BamHI+EcoRI restriction endonucleases and ligated into the LXSN vector. The complexity of the resulting libraries was approximately 30-40x10<sup>6</sup> individual clones. The size of the inserts and proportion of recombinant clones were checked by PCR

using sense and antisense primers corresponding to flanking vector sequences (Figure 12).

Fragmented retroviral libraries of similar characteristics were constructed from cDNAs for human p53, ATM, MDM2 and E6-AP genes. For example, the library of p53 fragments contained inserts between 250 and 1300 bp (total length of two adapters is 170 bp, total length of the full size of the coding region of p53 cDNA - 1200 bp). More than 95% of the clones contained inserts.

10

Example 2. Construction of tagged retroviral cDNA libraries

Two retroviral tagged libraries of mouse embryonic cDNA were constructed, using a mixture of poly(A) RNA isolated from 7, 11, 14, 17 day-old mouse embryos. Figure 6 shows the structure of oligonucleotide which was used as primers for cDNA synthesis for the first library. This oligonucleotide contains 18 dT residues in its 3' end, EcoRI and HindIII sites (to cut TAGs' arms), 80 bp random fragment (TAG itself) and SfiI(B) site for cloning. A library was constructed with Gibco BRL cDNA synthesis kit, 5' adapter contained ATG codon and SfiI(A) site for cloning. Such design of inserts allowed oriented cloning to be done using a single restriction enzyme. The library was cloned in LXSN SfiI retroviral vector (LXSN with modify MCS). Frequency and size of inserts was checked by PCR using sense GGCTAGGCCATTAAGGCCACC (specific for 5' adapter), and antisense GGATCCGGCCGAGGCAGGCC (specific for 3' TAGs "arm") primers. Library containing 97% of recombinant clones with the inserts ranging between 500-2000 bp (the total size 3' and 5' adapters was 170 bp). The total complexity of the library is approximately 1,000,000 individual clones. This library was prepared as independent separate batches containing 30-40 thousand clones in each. The below-

described experiments were done using two randomly picked batches.

Another library from the same starting RNA was constructed using similar procedures with the following 5 modifications introduced to improve the efficacy of recovery of tags and inserts. New tagging oligonucleotide was used containing 14 dT residues at its 3'-end, a new flanking sequence for efficient PCR amplification, the stretch of 80 random nucleotides flanked by EcoRI/HindIII 10 sites and SfiI (B) site for cloning. 5' adapter was identical to that used for the above-described library. The library was cloned in pLXSN vector with a modified sequence around cloning site (see Figure 7a,b). This modification involves introduction of double SfiI site 15 for directed cloning followed by T7 promoter in antisense orientation and 20 bp sequence termed Na1GT+ that is known as being highly efficient as a PCR primer (Luda Diatchenko, personal communication). This library has similar complexity and proportion of recombinant clones.

20

Example 3. New retroviral ("retrophage") vectors for generation and efficient screening of expression libraries

The phage lambda-based vectors have significant advantages over plasmid vectors in preparation and 25 screening of complex libraries. The efficacy of ligation of the inserts into phage vector significantly less depends on the size of insert (within the regular range of lengths of the majority of cDNAs). Phages can be plated at much higher density than bacterial colonies 30 thus increasing the efficiency of screening of large libraries. Hybridization screening of the prints of phage plaques on nylon or nitrocellulose is a far more sensitive and reliable technique than that of the bacterial colonies. Prints of plaques of phage libraries 35 are highly reproducible and can be considered cDNA arrays

allowing one to simultaneously analyze expression of numerous genes in a single hybridization.

At the same time, retroviral expression vectors have a series of advantages specifically valuable for 5 generation and screening of expression libraries (cDNA, GSE, antisense RNA, etc.). They represent the most efficient delivery vehicle that brings the gene of choice into the cell and determines its integration and stable expression in the genome at different levels within the 10 range of expression of cellular genes.

The advantages of phage and retroviral vectors were combined by creating a new retrovirus expression vector inside the phage DNA. The schematic presentation of this new vector is shown at the Figure 3. The new vector 15 consists of the λgt11 phage with LXSN retrovirus vector inside. The polylinker inside LXSN was substituted for two SfiI sites with different cohesive ends to make it possible oriented cloning of the inserts and to avoid self-ligation of the two arms of the phage DNA. λgt11 can 20 accept the fragments of DNA up to 7.2 kb in length. The size of provirus part is about 3 kb. Thus, the cloning capacity of the new vector is up to 4 kb.

To prove the efficacy of the created vector, its ability was directly compared to produce infectious 25 retroviruses after transfection into packaging cells with that of the LXSN plasmid vector and found it to be only 2 to 3 fold lower if calculated per microgram of DNA. The phage/retrovirus vector can be very useful for preparation of large expressing cDNA libraries with millions of 30 recombinant clones. To increase the cloning capacity of the phage/virus vector a new modification was made of the original vector, named λgtLXL, which lacks selectable marker gene and has cloning limit up to 5 kb.

Example 4. The use of tags to monitor negative selection of growth suppressive clones in retroviral library of fragmented p53

p53 is a key mediator of cell response to various types of stress. p53 gets activated after a variety of different stimuli (DNA damage, block of DNA replication, disruption of microtubules, cell detachment, hypoxia, etc.) and determines cell cycle arrest or apoptosis. To isolate the truncated growth suppressive version of p53 SSA was applied to screen the library of randomly fragmented p53 cDNA. The library was constructed as described above. Briefly, human p53 cDNA was partially digested by DNase I in the presence of Mn<sup>++</sup> and the cDNA fragments ranging in size between 200 bp and full-length cDNA were ligated with two synthetic adaptors. One adaptor ("left") contained ATG codon in Kozak's consensus and EcoRI restriction site. A second adaptor ("right") contained stop codons in all reading frames and 80 bp random sequence ("tag") flanked by two 20-bp specific sequences for PCR amplification and BamHI restriction site (see Figures 4 and 12). After ligation with adaptors, the ligation mixture was digested by EcoRI and BamHI and cloned into a PLXSN retroviral vector expressing the inserts from the MoMuLV LTR promoter. The library of 30,000 independent clones was generated (Figure 13). The library was converted into a mixture of retroviral particles by transfection into BOSC23 packaging cells and the virus was used to infect several types of cells known to be sensitive to p53: HeLa cells (cervical carcinoma) expressing ecotropic viral receptor, mouse and human embryonic fibroblasts deficient in p53 (lines 10(1) and 041, respectively) and mouse embryonic fibroblasts transformed with Ela+ras (C8 cells) and highly sensitive to p53-dependent apoptosis. RNA was isolated (RNA pool 1) from each cell type 48 hours after infection and the remaining cells (50% of infected

population) were maintained for one week after which cells were harvested for RNA isolation (RNA pool 2). RT PCR was performed on each RNA sample using primers for the sequences flanking 80-bp tags (the exact sequences of 5 adaptors are shown in Figure 4). The resulting PCR products were used as a tester (tag pool 1) and driver (tag pool 2) DNAs for SSH subtraction procedure (Figure 10).

To insure that the loss of "toxic" clones occurs 10 during selection, we isolated a clone from the library representing full-length p53 cDNA and used its tag as a probe for Southern hybridization to trace its behavior during selection. As it is shown in Figure 14, the hybridization signal of this clone is strongly decreased 15 in both selected populations of target cells. These observations demonstrate that random 80 bp tag sequences can be used to monitor elimination of growth suppressive clones in tagged retroviral library delivered to target cells by hybridization.

20

Example 5. Screening of tagged retroviral library of p53 cDNA fragments for growth suppressive clones

To imitate a microarray hybridization-based screening for differentially represented tags, a 25 sublibrary of randomly picked 2,000 clones was separated from the tagged p53 library (see Example 4), all clones of which were printed onto nylon to form an array of clones for hybridization screening (see Figure 15). This mini-library was delivered to p53-sensitive 10(1) cells 30 and tags were isolated by RT-PCR as described in Example 4 from cellular RNA isolated 24 hours after retrovirus infection and after completion of G418 selection (8 days after infection). Both isolates of tags were labeled with <sup>32</sup>P using a primer extension by Klenow fragment (tag 35 adaptor sequences were used as primers) and the probes were hybridized separately with the library-carrying

nylon membrane. Clones that demonstrated differential hybridization signal (low signal after G418 selection) were picked, grown and mixed together with several randomly picked control clones that did not show any changes after selection. This mixture was subjected to a second round of selection, tag recovery and hybridization. As it is shown in Figure 15, the majority of the clones originally picked for differential hybridization were confirmed in secondary screening.

10 Seven such clones were isolated whose biological activity was individually confirmed by a direct transduction of target cells. p53 inserts were sequenced and were all found to contain full-length p53 cDNA. It is noteworthy that the proportion of full-length cDNA in the library

15 was lower than 1% as judged by the results of PCR on individual randomly picked library clones.

These results indicate that identification and isolation of growth suppressive clones based on tag differential hybridization procedure can be done

20 efficiently.

Example 6. Monitoring the elimination of growth suppressive clones from cDNA retroviral library in different cell types

25 SSA methodology involves delivery of expression library into target cells, isolation of the inserts from the cells shortly after delivery (before selection) and after selection and identification of sequences that are differentially represented in two DNA populations

30 recovered according to above described methods. However, the whole strategy would work only if the selection procedure applied is in fact sufficient to cause significant changes in the proportion of certain clones in the delivered library. To address this question,

35 several retroviral cDNA libraries containing cDNA sequences encoding some known growth suppressors were

delivered in several cell lines in order to set up the conditions of insert rescue and to monitor the fate of growth suppressive inserts during cell propagation. The results of these experiments are illustrated by Figures 5 19-21. Retroviral libraries of full-length cDNAs from HeLa cells were used, human placenta and normal human mammary gland manufactured by CLONTECH Laboratories. Retroviral vector pLIB used for the construction of these libraries contains no additional promoters or selectable 10 markers to increase its cloning capacity. These libraries were converted from plasmid forms into mixtures of retroviruses by transfection of either ecotropic (BOSC23) or amphotropic (CAK8) packaging cells (Pear et al., 1993) and the resulting viruses were used to infect target 15 cells. As target cells, human breast carcinoma cells MCF7 were used, human cervical carcinoma cell line HeLa and mouse embryo fibroblasts transformed by a combination of Ela and activated ras, line C8. Total RNA was isolated from part of the infected cells 24 hours after infection, 20 while the rest of the infected population was maintained for additional 10 days and RNA was isolated at each passage of the infected cultures (1:3). cDNA was synthesized on all RNA preparations, using either vector-specific or random primers for reverse transcriptase. 25 Library inserts were amplified from these cDNA samples by PCR with the primers corresponding to vector sequences flanking the inserts. Large number of primer combinations were tested to find the optimal pair that would allow to specifically amplify the inserts with no or low 30 background on control cDNA synthesized using RNA from non-infected cells. The following pair of primers gave acceptable results (Figure 19) and therefore was chosen for further applications:

Sense primer (LXLS4): 5'-CCT TGA ACC TCC TCG TTC G-3'

Antisense primer (reverse 3) 5'-TTA CTT AAG CTA GCT TGC  
CAA ACC-3'.

PCR products consisting of the recovered library  
5 inserts are shown in Figure 20. They were separated in  
agarose gels and analyzed by Southern hybridization. As  
probes for hybridization, the sequences encoding either  
non-toxic, non-cytostatic protein (ubiquitin) or known  
growth inhibitory protein (p21/WAF1) were used to monitor  
10 the presence and the proportion of clones with different  
biological activity during culture growth. The results of  
this analysis are shown in Figure 21. While the intensity  
of hybridization of ubiquitin probe does not change in  
passages, the proportion of p21/WAF1 sequences is  
15 significantly reduced consistently with the known growth  
 suppressive effect of p21/WAF1 overexpression.

Example 7. Isolation of tagged growth suppressive clones  
using RNase H-based subtraction of tags

20 To set up the conditions allowing one to isolate  
growth suppressive clones from a retroviral expression  
library using tag-based subtraction selection approach  
(SSA) a tagged retroviral library of mouse embryonic cDNA  
cloned in modified pLXSN vector was used (see library  
25 description above). In order to be able to monitor the  
efficacy of the procedure of isolation of differentially  
represented tags, this library was mixed with a small  
amount (1/1,000) of a retroviral vector carrying  
poliovirus gene 2A tagged with a known 80 bp tag  
30 sequence. This construct had exactly the same design as  
any library-derived clone. Polioviral gene 2A encodes  
highly growth suppressive protein that is known to act as  
a potent inhibitor of cap-dependent translation. Strong  
biological effect of this vector was confirmed in gene  
35 transfer experiments (retrovirus transduction or

transfection). A representative result of such testing is presented in Table 1.

Table 1. G418 resistant colony formation by the indicated target cells infected or transfected by the indicated vectors. Numbers in parenthesis correspond to the proportion of positive cells during transient phase of transfection (monitoring by co-transfection with the GFP-expressing vector).

Construct	Delivery	Cells		
		HeLa	10(1)	Rat1
pLXSN (empty vector)	Transfection	5,000 (20%)	19,000 (12%)	12,000 (17%)
	Infection (titer)	110,000	180,000	210,000
pL(2A)SN (carrying 2A gene)	Transfection	0 (22%)	0 (12%)	4 (15%)
	Infection (titer)	0	0	10

10

Tagged library including 0.1% of 2A-carrying clones was delivered to three different cell lines (Rat1 cells; Balb 3T3 cells, line 10(1) and human cervical carcinoma HeLa cells) by retrovirus transduction as described above. Cells were selected on G418. RNA was isolated from the cells 24 hours after library delivery and after G418 selection is completed and used for the RNase H-based subtraction procedure. Tag sequences were isolated by RT-PCR from the post-selection pool of RNA. Universal primer-derived sequences were removed from tags by restriction digestion followed by gel purification.

The resulting tag DNA was denatured and annealed with RNA prepared from the cells taken prior to selection as described in Methods section. After treatment with RNase H, the remaining RNA was used for RT-PCR with vector-derived primers flanking tag sequences and with the primers allowing isolation of the complete insert (Figure 17). The presence of the specific tag, marking 2A-carrying clones, was determined by Southern hybridization in the original and subtracted tag populations (Figure 22). The obtained results show that RNase H-based procedure resulted in a significant enrichment of the tags corresponding to the cytotoxic insert.

Throughout this application, various publications, are referenced by author and year. Full citations for the publications are listed below. The disclosures of these publications in their entireties are hereby incorporated by reference into this application in order to more fully describe the state of the art to which this invention pertains.

The invention has been described in an illustrative manner, and it is to be understood that the terminology which has been used is intended to be in the nature of words of description rather than of limitation.

Obviously, many modifications and variations of the present invention are possible in light of the above teachings. It is, therefore, to be understood that within the scope of the appended claims, the invention may be practiced otherwise than as specifically described.

30

## REFERENCES

5 Burke and Olson, "Preparation of Clone Libraries in Yeast Artificial-Chromosome Vectors" in Methods in Enzymology, Vol. 194, "Guide to Yeast Genetics and Molecular Biology", eds. C. Guthrie and G. Fink, Academic Press, Inc., Chap. 17, pp. 251-270 (1991).

10 Capecchi, "Altering the genome by homologous recombination" Science 244:1288-1292 (1989).

15 Cregg JM, Vedvick TS, Raschke WC: Recent Advances in the Expression of Foreign Genes in *Pichia pastoris*, Bio/Technology 11:905-910, 1993

20 Davies et al., "Targeted alterations in yeast artificial chromosomes for inter-species gene transfer", Nucleic Acids Research, Vol. 20, No. 11, pp. 2693-2698 (1992).

25 Diatchenko L, Lau YF, Campbell AP, Chenchik A, Moqadam F, Huang B, Lukyanov S, Lukyanov K, Gurskaya N, Sverdlov ED, Siebert PD (1996). Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. Proc Natl Acad Sci U S A 93:6025-6030.

Dickinson et al., "High frequency gene targeting using insertional vectors", Human Molecular Genetics, Vol. 2, No. 8, pp. 1299-1302 (1993).

30 Duff and Lincoln, "Insertion of a pathogenic mutation into a yeast artificial chromosome containing the human APP gene and expression in ES cells", Research Advances in Alzheimer's Disease and Related Disorders, 1995.

35 Gilboa, E, Eglitis, MA, Kantoff, PW, Anderson, WF: Transfer and expression of cloned genes using retroviral vectors. BioTechniques 4(6):504-512, 1986.

40 Gudkov, A.V., Kazarov A.R., Thimmapaya, R., Mazo, I.A., Axenovich, S., and Roninson, I.B. (1994). Isolation of genetic suppressor elements from a retroviral normalized cDNA library: identification of a kinesin associated with drug sensitivity and senescence. Proc. Natl. Acad. Sci. USA 91, 3744-3748.

45

Gudkov, A.V. and Roninson, I.B. (1996). Isolation of genetic suppressor elements (GSEs) from random fragment cDNA libraries in retroviral vectors. In: *Methods in Molecular Biology*, Vol. 32: *Protocols for cDNA Libraries*.  
5 Humana Press, Inc. Edited by: Ian Cowell & Caroline Austin. P. 221-240.

Huston et al, 1991 "Protein engineering of single-chain Fv analogs and fusion proteins" in *Methods in Enzymology* (JJ Langone, ed.; Academic Press, New York, NY) 203:46-88.  
10

Huxley et al., "The human HPRT gene on a yeast artificial chromosome is functional when transferred to mouse cells by cell fusion", Genomics, 9:742-750 (1991).  
15

Jakobovits et al., "Germ-line transmission and expression of a human-derived yeast artificial chromosome", Nature, Vol. 362, pp. 255-261 (1993).  
20

Johnson and Bird, 1991 "Construction of single-chain Fvb derivatives of monoclonal antibodies and their production in *Escherichia coli*" in *Methods in Enzymology* (JJ Langone, ed.; Academic Press, New York, NY) 203:88-99.  
25

Lamb et al., "Introduction and expression of the 400 kilobase precursor amyloid protein gene in transgenic mice", Nature Genetics, Vol. 5, pp. 22-29 (1993).  
30

Lisitsyn N, Lisitsyn N, Wigler M (1993). Cloning the differences between two complex genomes. *Science* Feb 12;259(5097):946-951.

Mernaugh and Mernaugh, 1995 "An overview of phage-displayed recombinant antibodies" in *Molecular Methods In Plant Pathology* (RP Singh and US Singh, eds.; CRC Press Inc., Boca Raton, FL) pp. 359-365.  
35

Pearson and Choi, Expression of the human b-amyloid precursor protein gene from a heast artificial chromosome in transgenic mice. *Proc. Natl. Scad. Sci. USA*, 1993. 90:10578-82.  
40

Rothstein, "Targeting, disruption, replacement, and allele rescue: integrative DNA transformation in yeast" in Methods in Enzymology, Vol. 194, "Guide to Yeast Genetics and Molecular Biology", eds. C. Guthrie and G. Fink, Academic Press, Inc., Chap. 19, pp. 281-301 (1991).  
45

Russ AP, Friedel C, Grez M, von Melchner H (1996). Self-deleting retrovirus vectors for gene therapy. *J Virol* 70:4927-4932.

- 5 Schedl et al., "A yeast artificial chromosome covering the tyrosinase gene confers copy number-dependent expression in transgenic mice", Nature, Vol. 362, pp. 258-261 (1993).
- 10 Strauss et al., "Germ line transmission of a yeast artificial chromosome spanning the murine  $\alpha_1$  (I) collagen locus", Science, Vol. 259, pp. 1904-1907 (1993).

What Is Claimed:

1. A method of selection subtraction comprising the steps of:

5 creating an expression library wherein each clone is marked with a unique tag;  
delivering the vector to target cells;  
subjecting cells to a selection procedure; and  
recovery of tags from cell before and after selection; and  
10 comparing differentially represented tags and using the tags to identify the desired clones which represent the desired phenotypes.

15 2. The method according to claim 1 wherein said method further includes screening a sample for the tagged clones.

20 3. An expression library according to claim 1 comprising individual clones marked with a unique tag sequence.

4. An expression library according to claim 1 wherein said clones include mixtures of fragments of genomic DNA or cDNA.

25

5. An expression library according to claim 1 wherein said clones include mixtures of artificial sequences.

6. The method according to claim 1 wherein  
the comparison of tagged clones comprises isolating  
genetic elements associated with complete or partial cell  
growth suppression or cell death, by isolating from the  
**5** tagged expression library those clones that are lost in  
the cell population.

7. The method according to claim 6, wherein  
said isolating step further includes rescuing the clone  
**10** from the cell before and after selection.

8. The method according to claim 7 wherein  
said rescuing step further includes rescuing using a  
method from the group consisting essentially of PCR, RT-  
**15** PCR, direct transformation, virus rescue, or gene  
subtraction techniques.

9. A method of creating expression libraries  
by marking all clones with unique tags.  
**20**

10. The method according to claim 9, wherein  
said method further includes screening a sample for the  
tagged clones.

**25** 11. An expression library comprising  
individual clones marked with a unique tag sequence.

12. The expression library according to claim  
11, wherein said claim includes any expressible sequence.  
**30**

13. The expression library according to claim  
11, wherein said clone includes mixtures of cDNAs cloned  
in sense and anti-sense orientation.

14. The expression library according to claim 11, wherein said clone includes mixtures of fragments of genomic DNA or cDNA.

5 15. The expression library according to claim 11, wherein said clone includes mixtures of artificial sequences.

10 16. A method of isolating genetic elements which induce any cell phenotype by isolating clones from an expression library which exhibit the selective trait.

15 17. The method according to claim 16, wherein said isolating step further includes isolating clones from an expression library which are associated with complete or partial growth suppression, by isolating from the expression library those clones that are lost in the cell population.

20 18. The method according to claim 17, wherein said isolating step further includes rescuing the clone from the cell before and after selection.

25 19. The method according to claim 18, wherein said rescuing step further includes rescuing using a method from the group consisting essentially of PCR, RT-PCR, direct transformation, virus rescue, or gene subtraction techniques.

30 20. A method of identification of differentially represented sequences in two populations of nucleic acids wherein one is represented by RNA and another with DNA comprising the steps of:

35 annealing RNA and DNA sequences;  
treatment with RNaseH or similar enzyme destroying RNA within RNA/DNA hybrid;

copying and amplification of remaining RNA molecule by polymerase reaction primed from sequences located outside of the specific RNA sequence.

5           21. The method according to claim 20 wherein said method further includes screening a sample for the tagged clones.

10          22. A method of subtraction comprising the steps of:

subjecting cells to an RNaseH-based subtraction procedure;

recovering tags or inserts from cells before and after selection;

15          comparing differentially represented tags or inserts; and

using the tags or inserts to identify the desired clones which represent the desired phenotypes.

20          23. A method of selection subtraction comprising the steps of:

creating an expression library with multiple clones;

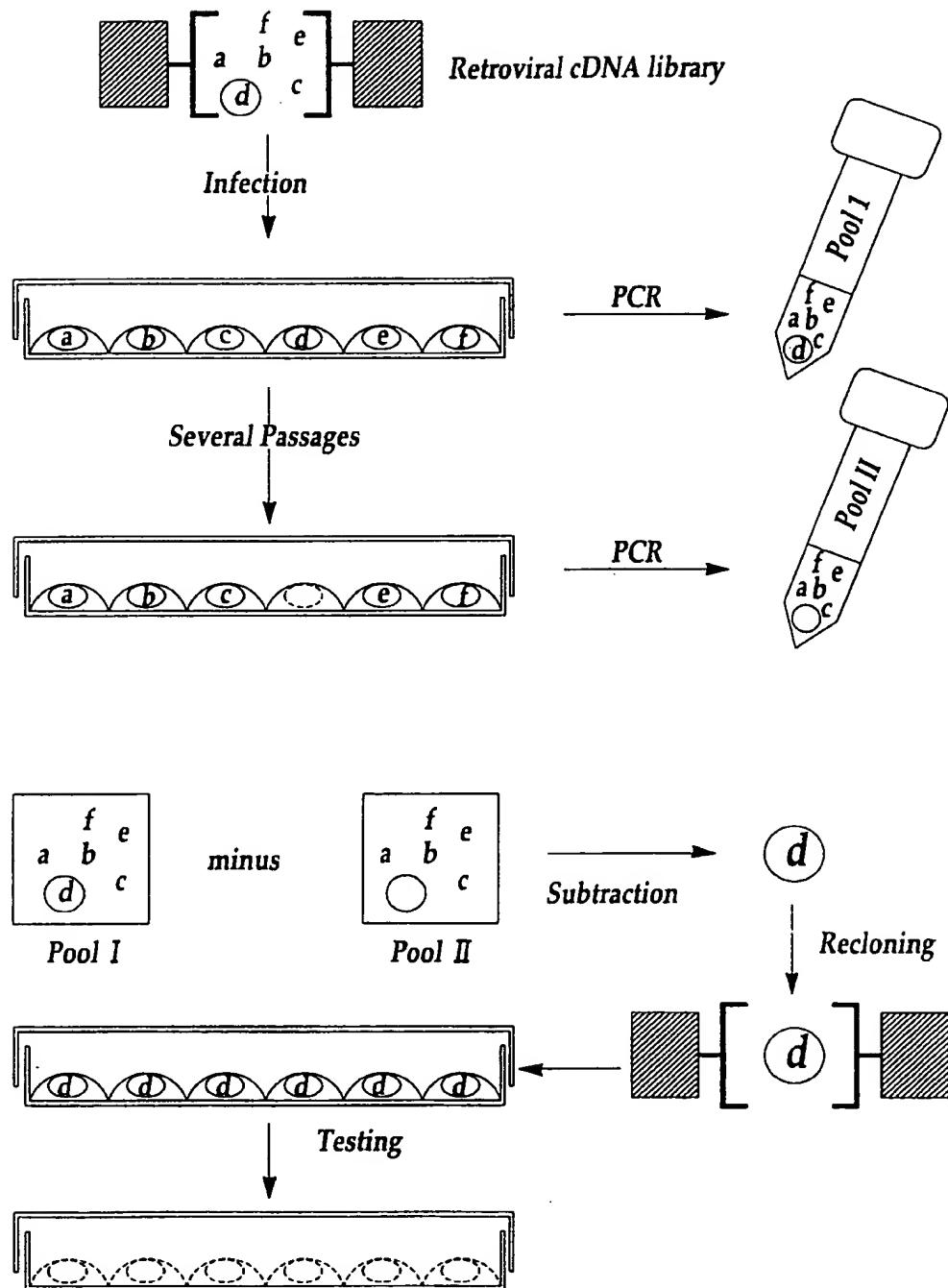
delivering the vector to target cells;

25          subjecting cells to a selection procedure; and

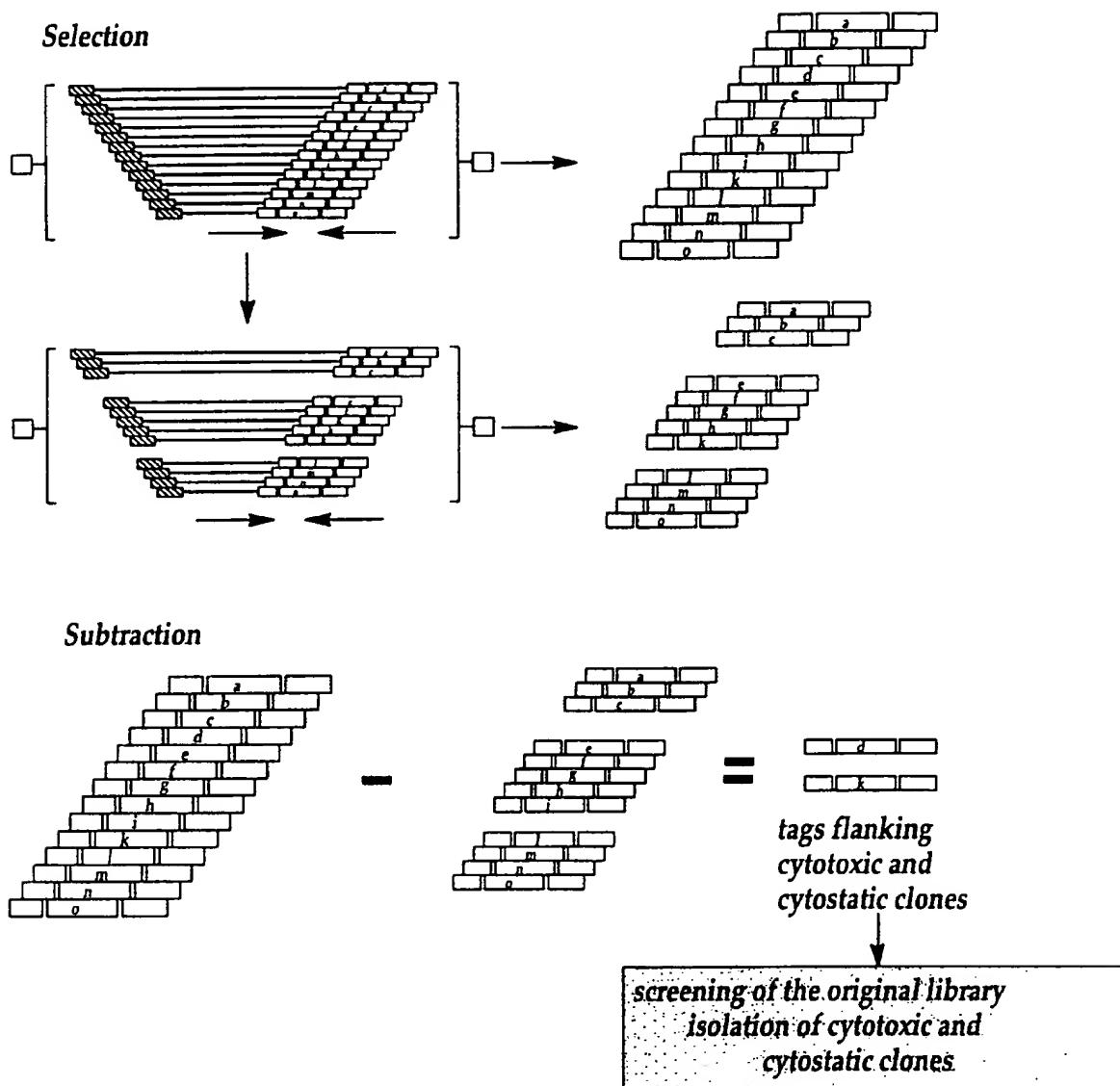
recovery of inserts from cell before and after selection; and

30          comparing differentially represented inserts and using the inserts to identify the desired clones which represent the desired phenotypes.

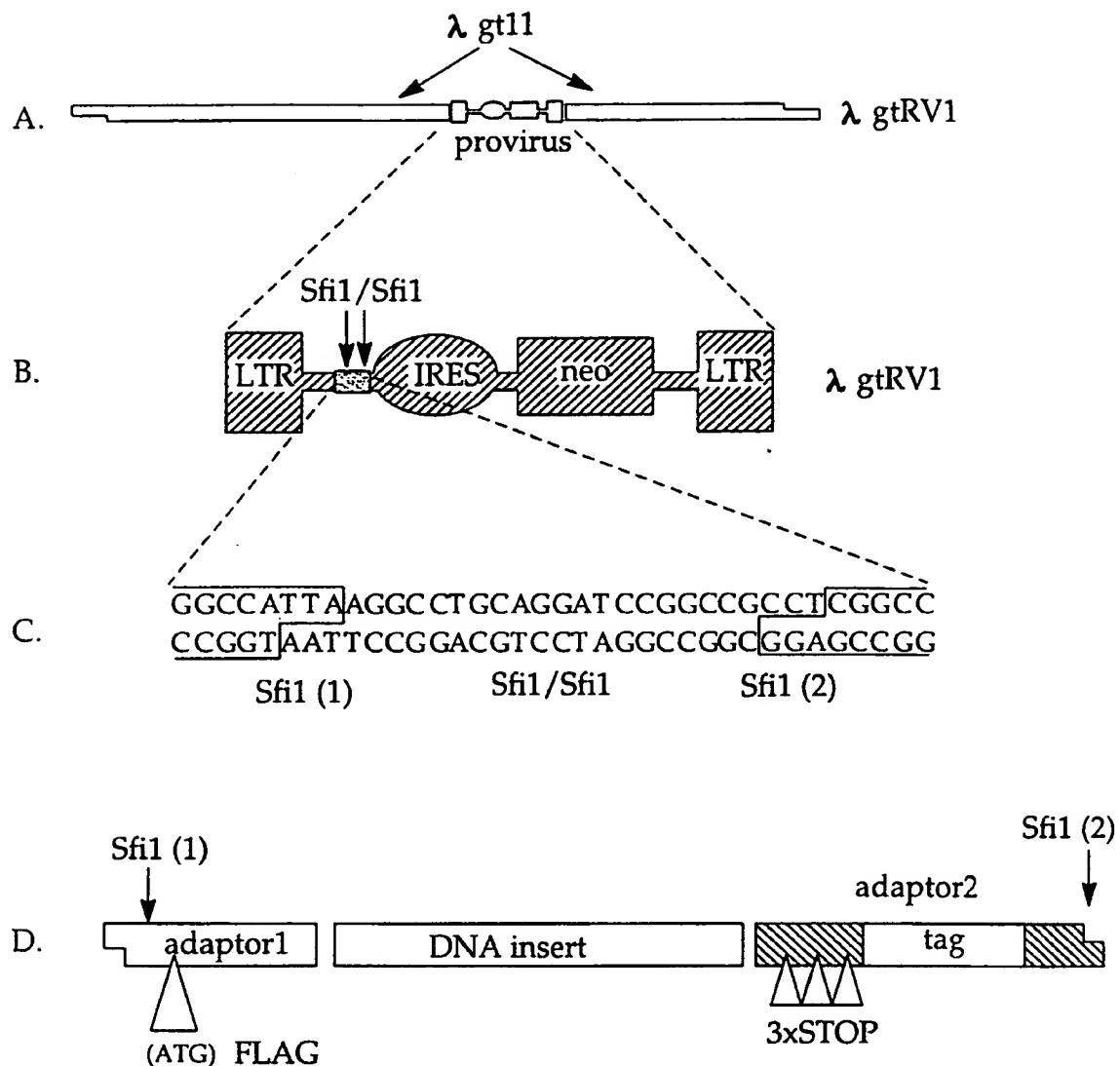
1/23

Fig-1

2/23

Fig-2

3/23

Fig-3

4/23

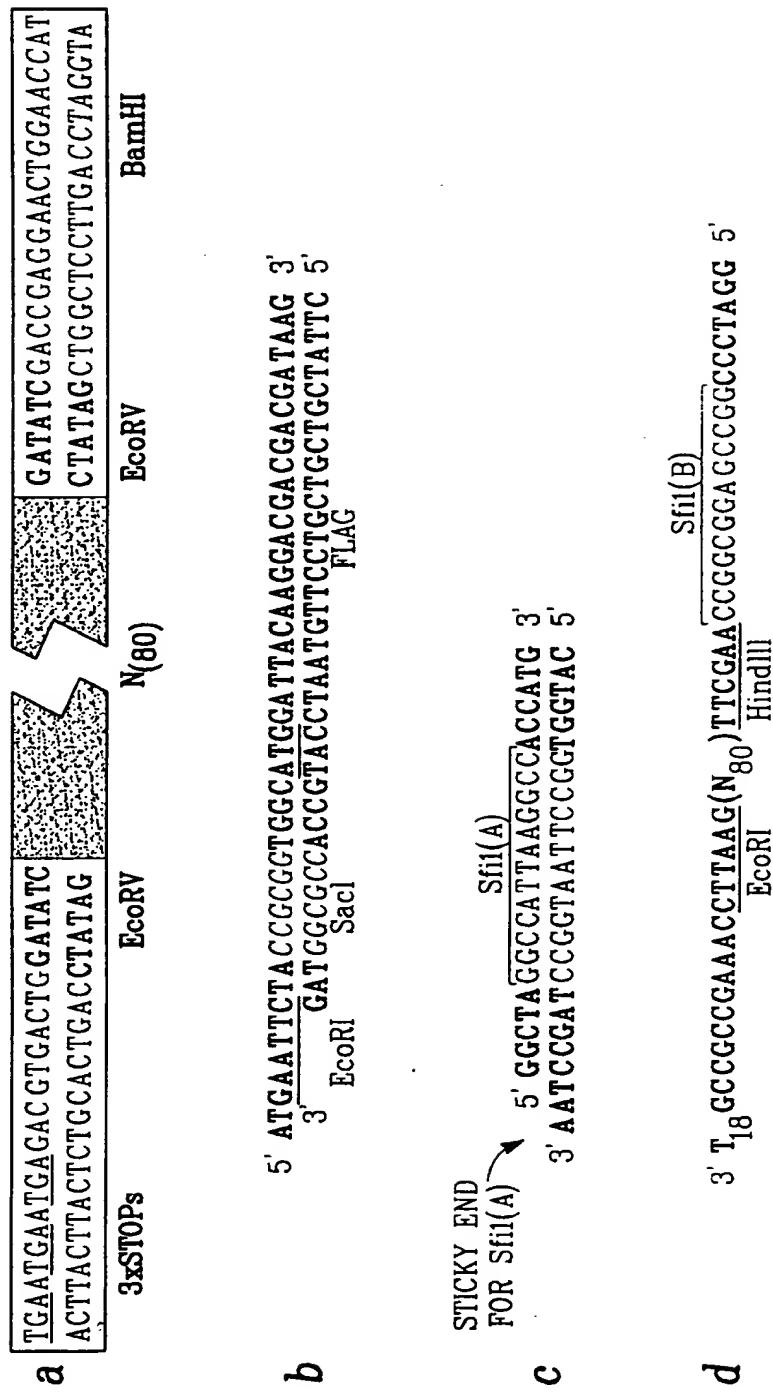
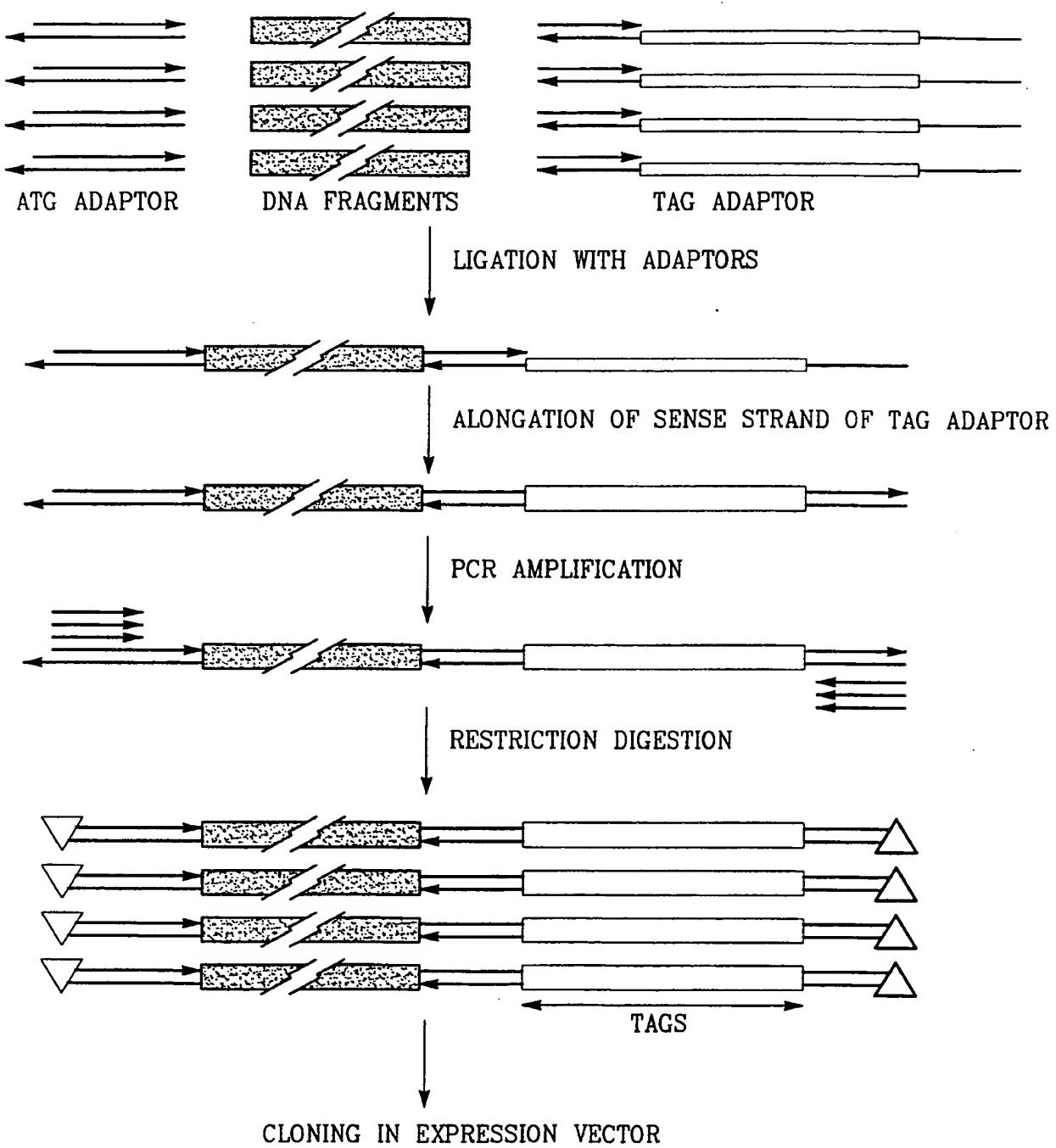
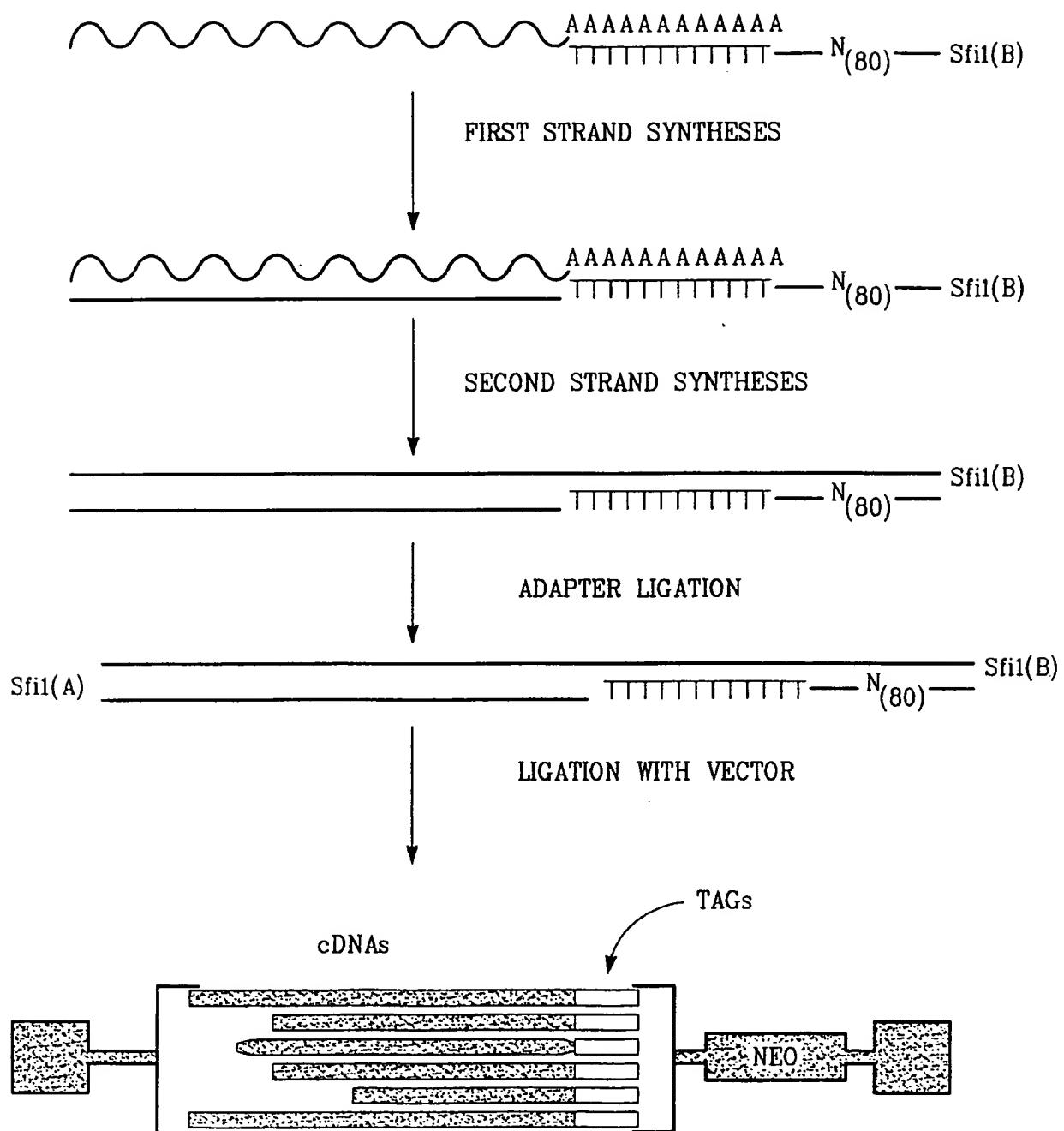


Fig-4

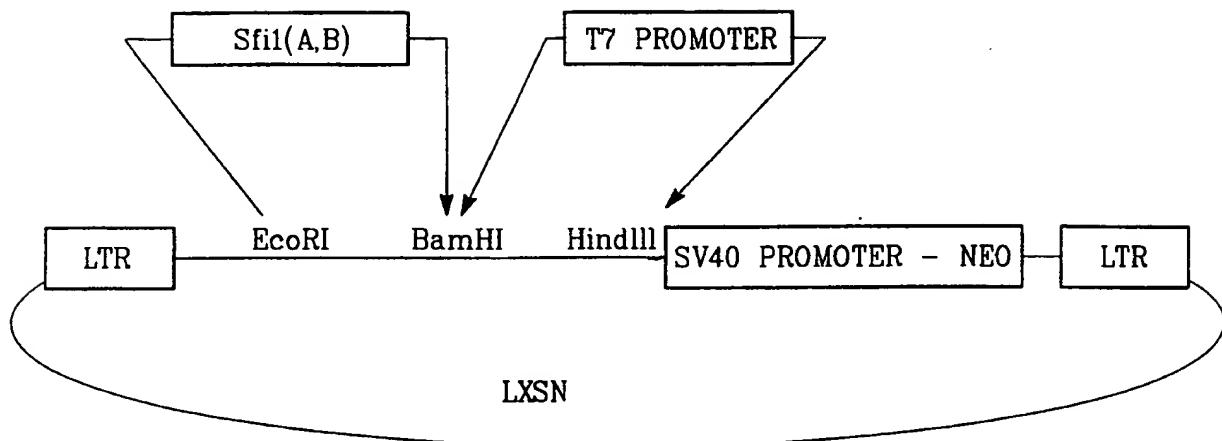
5/23

*Fig-5*

6/23

Fig-6

7/23



**STICKY END  
FOR BamHI**

## T7 PROMOTER NA1sGT+

**STICKY END  
FOR Hind III**

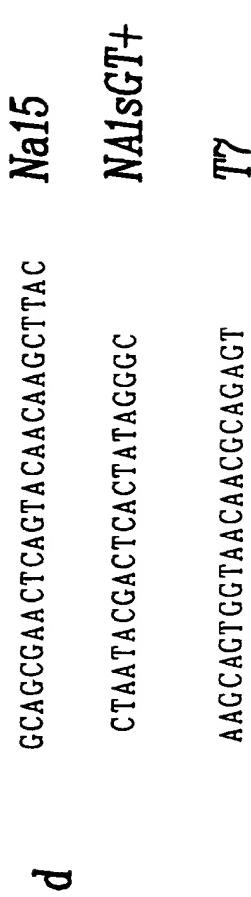
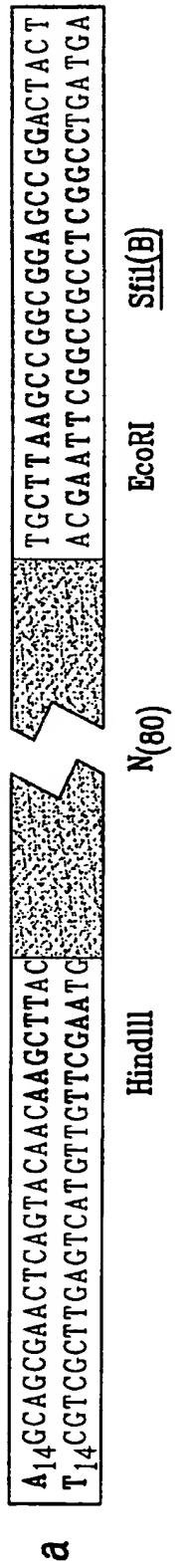
**STICKY END  
FOR BamHI**

**SfiI(A)** SfiI(B)  
AATTGGCCATTAAGGCCTGCAGGATCCGGCGCCTCGGCC  
CCGGTAATTCCGGACGTCTAGGGCGGCGAGCCGGCTAG

**STICKY END  
FOR BamHI**

Fig-7a

8/23

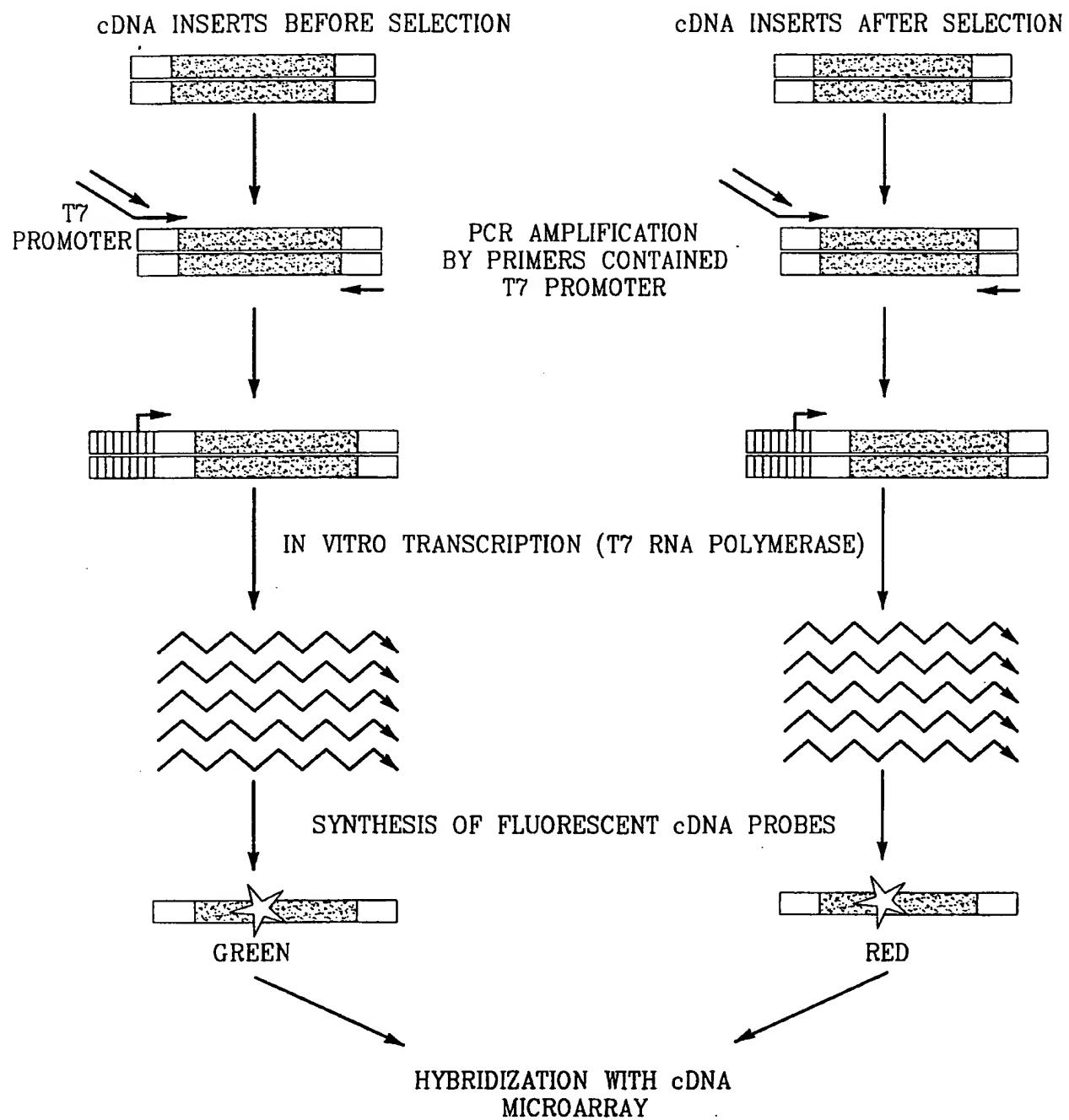


SUBSTITUTE SHEET (RULE 26)

Fig-7b

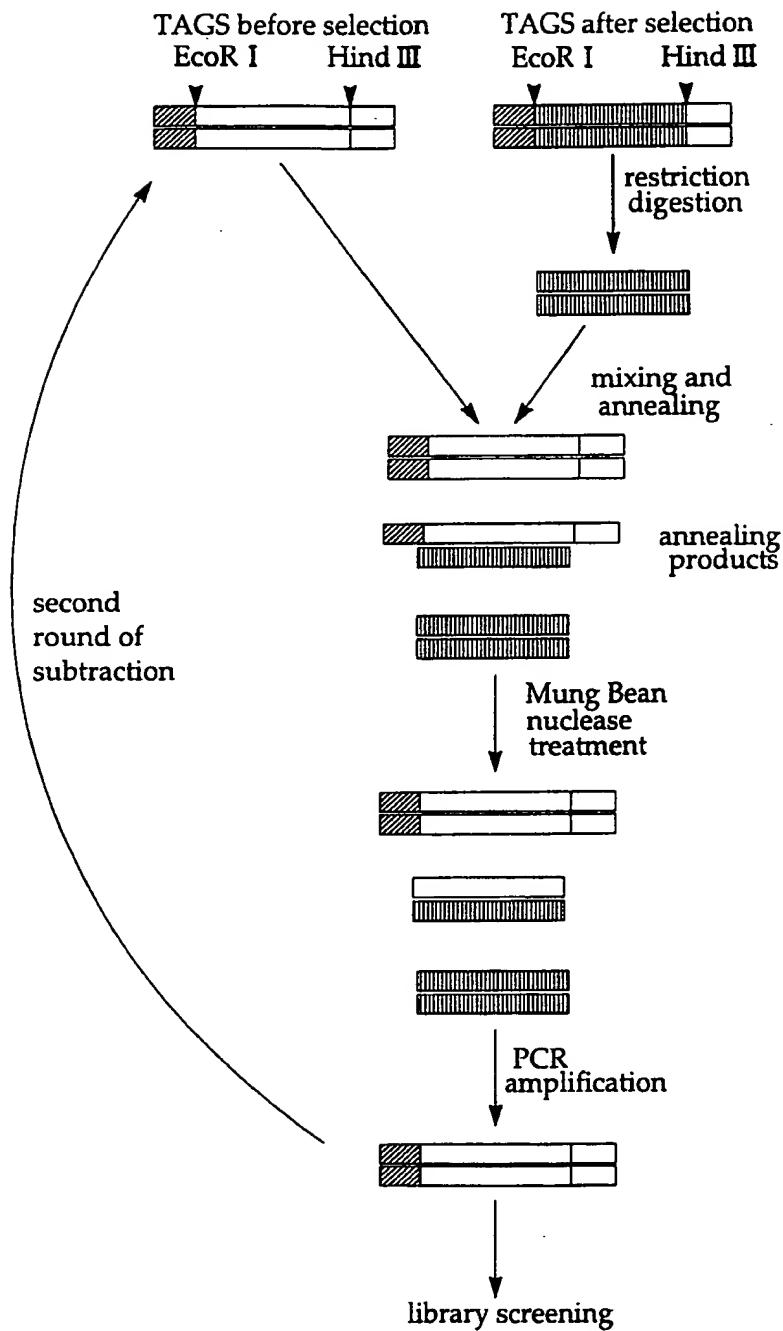
NEW DESIGN OF FULL SIZE TAGGED cDNA LIBRARY  
 a - STRUCTURE OF TAG, b - STRUCTURE OF 5' ADAPTER, c - STRUCTURE OF  
 3' OLIGONUCLEOTIDE USED FOR FIRST STRAND SYNTHESIS,  
 d - PRIMERS FOR TAG RESCUE/AMPLIFICATION

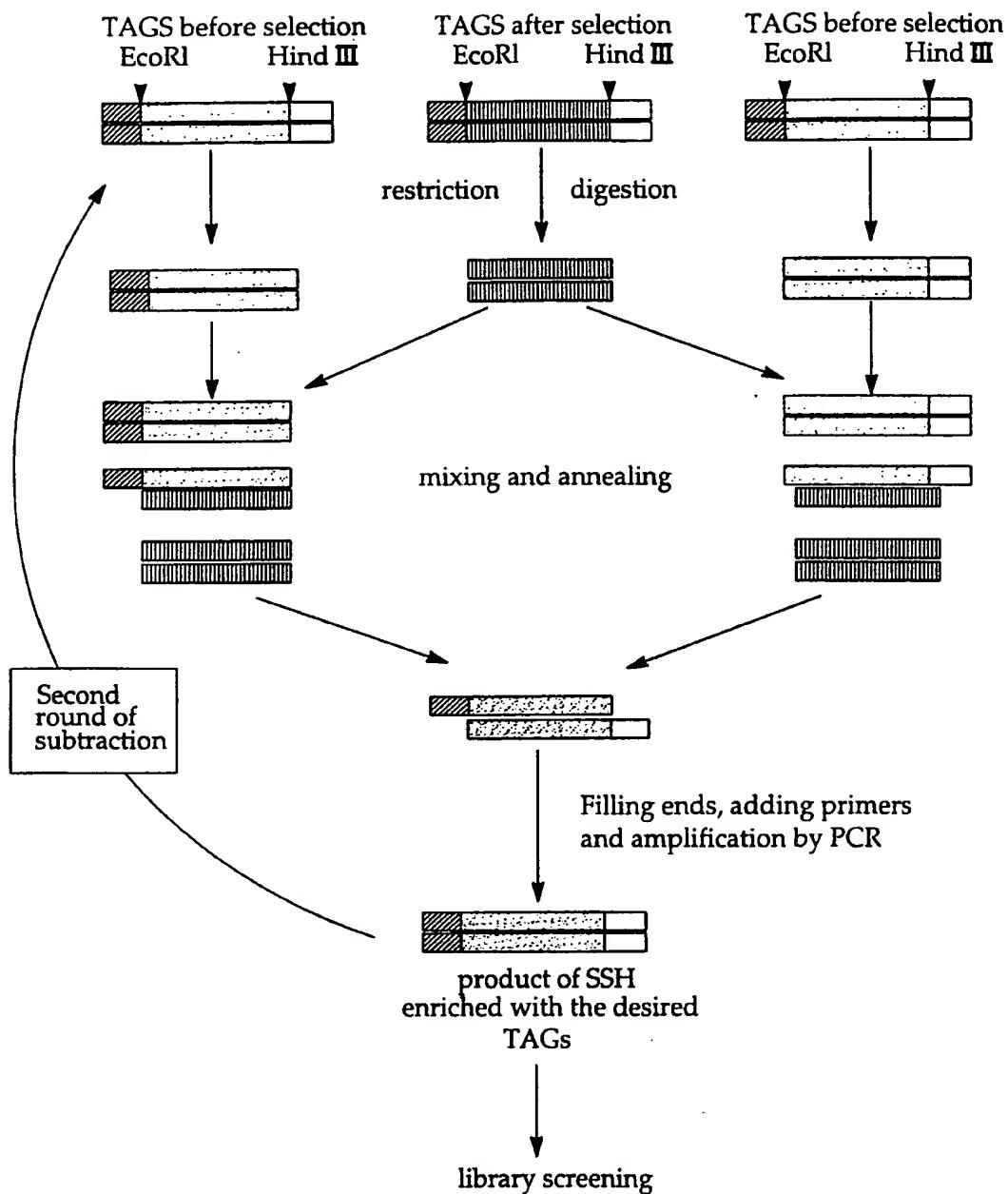
9/23

*Fig-8*

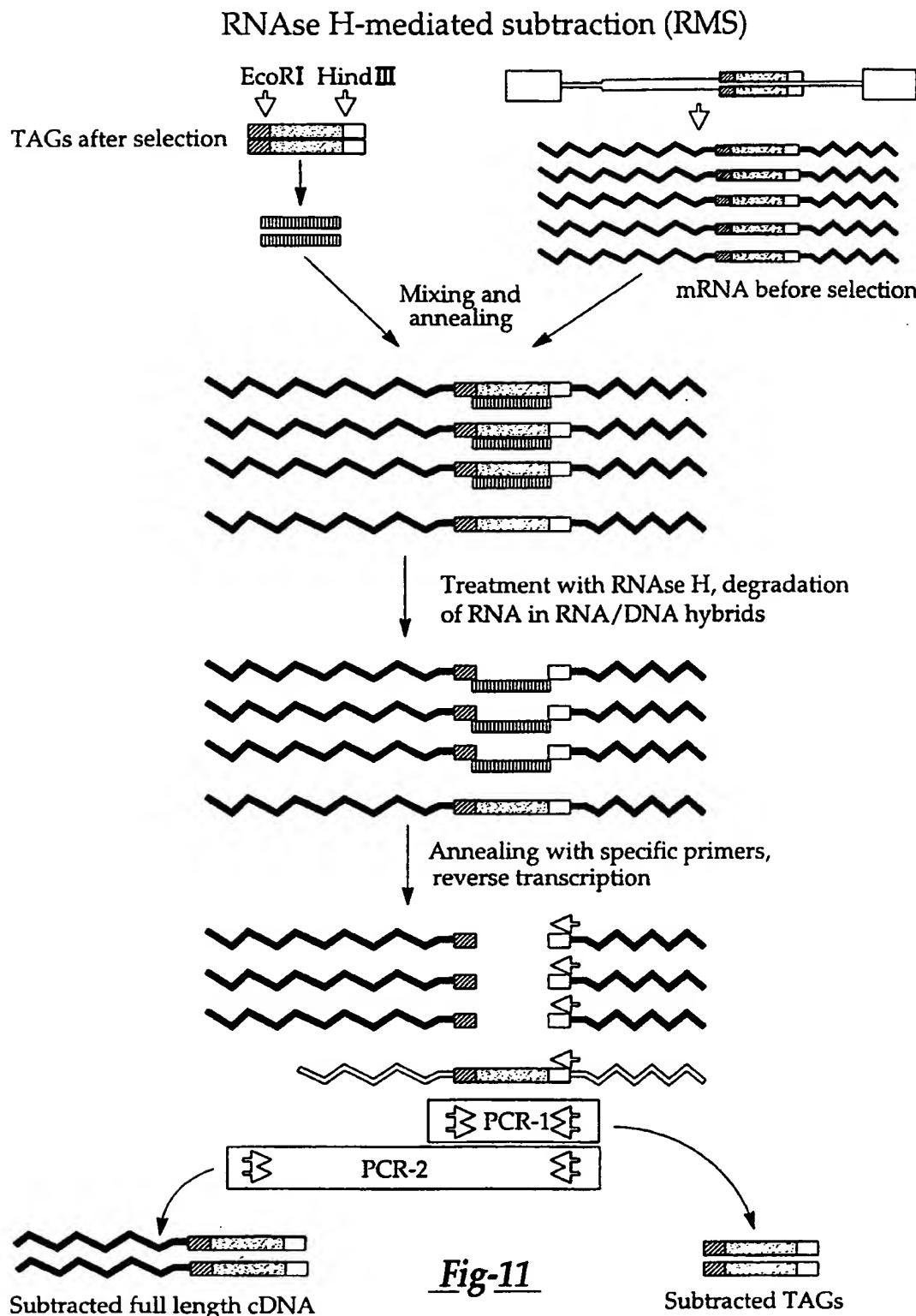
SUBSTITUTE SHEET (RULE 26)

10/23

Fig-9

*11/23**Fig-10*

12 / 23



13/23

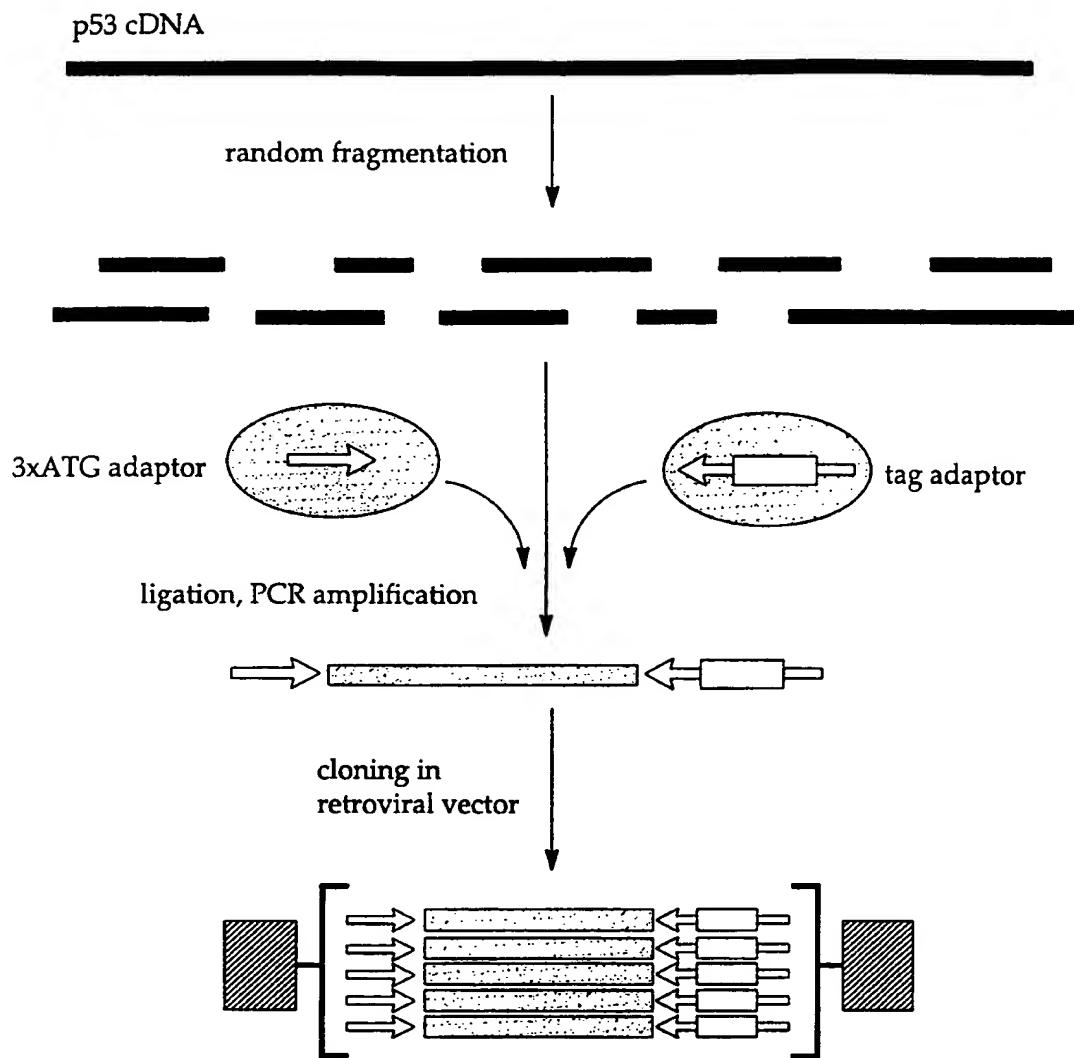
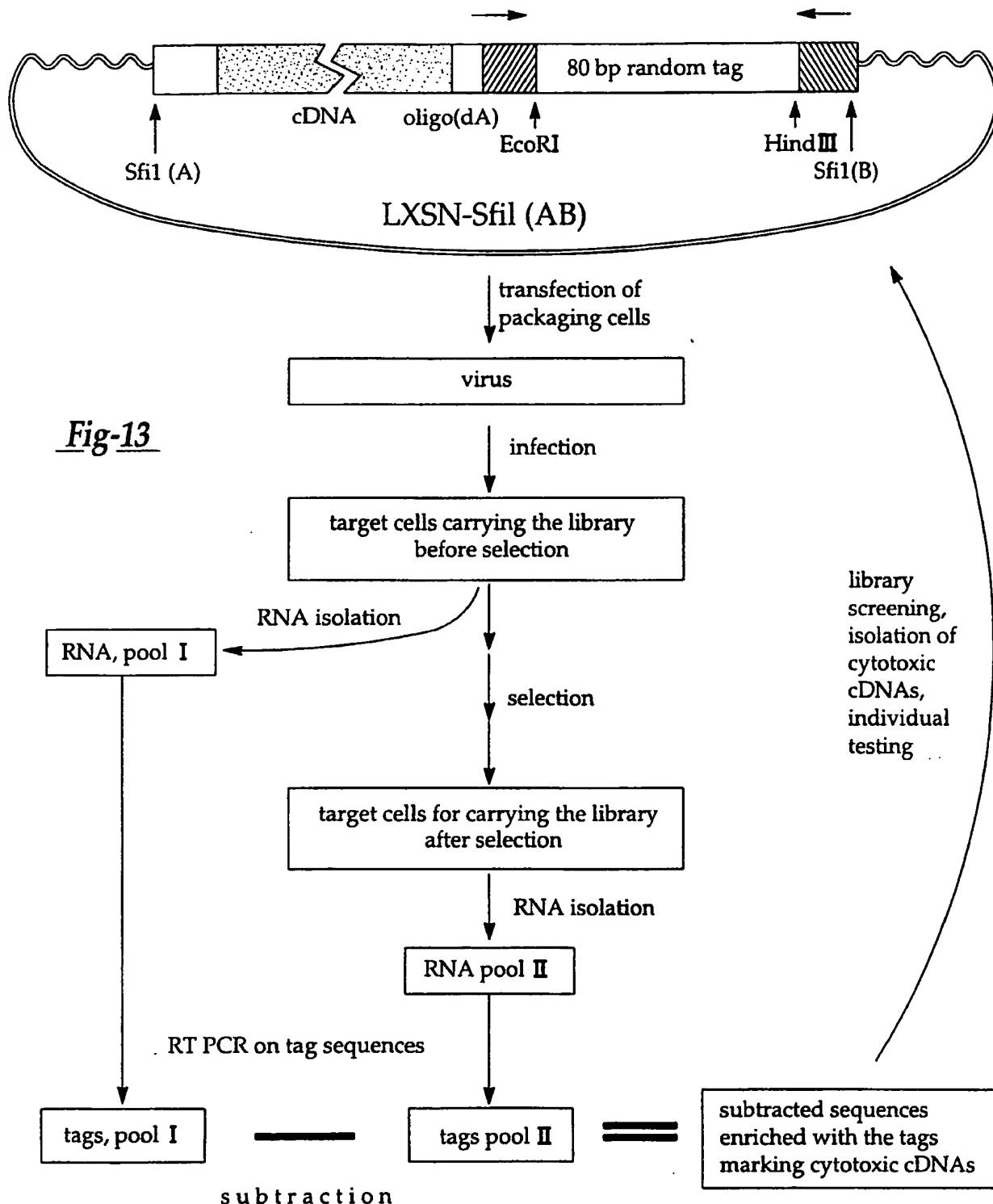
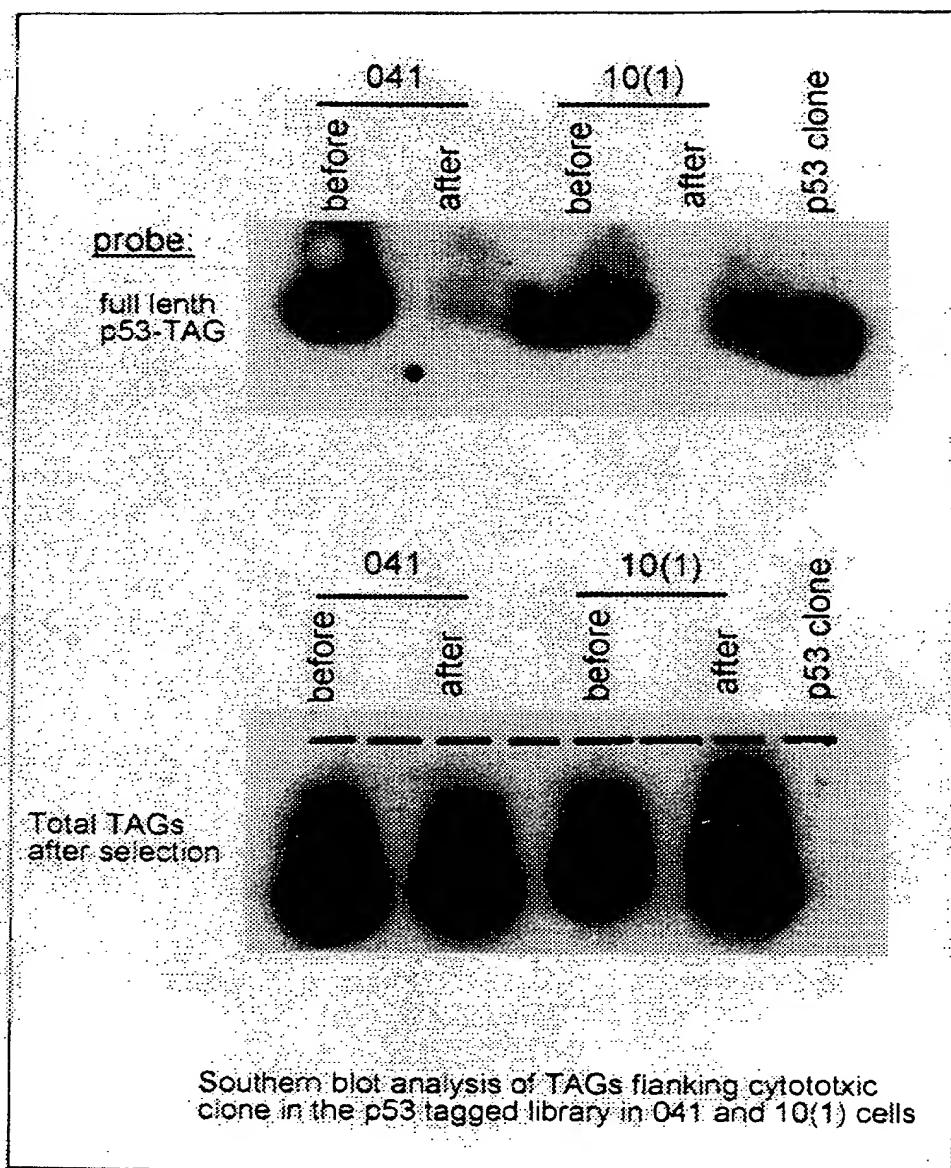


Fig-12

14/23



15/23

Fig-14

# BEST AVAILABLE COPY

WO 99/47643

PCT/US99/05944

16/23

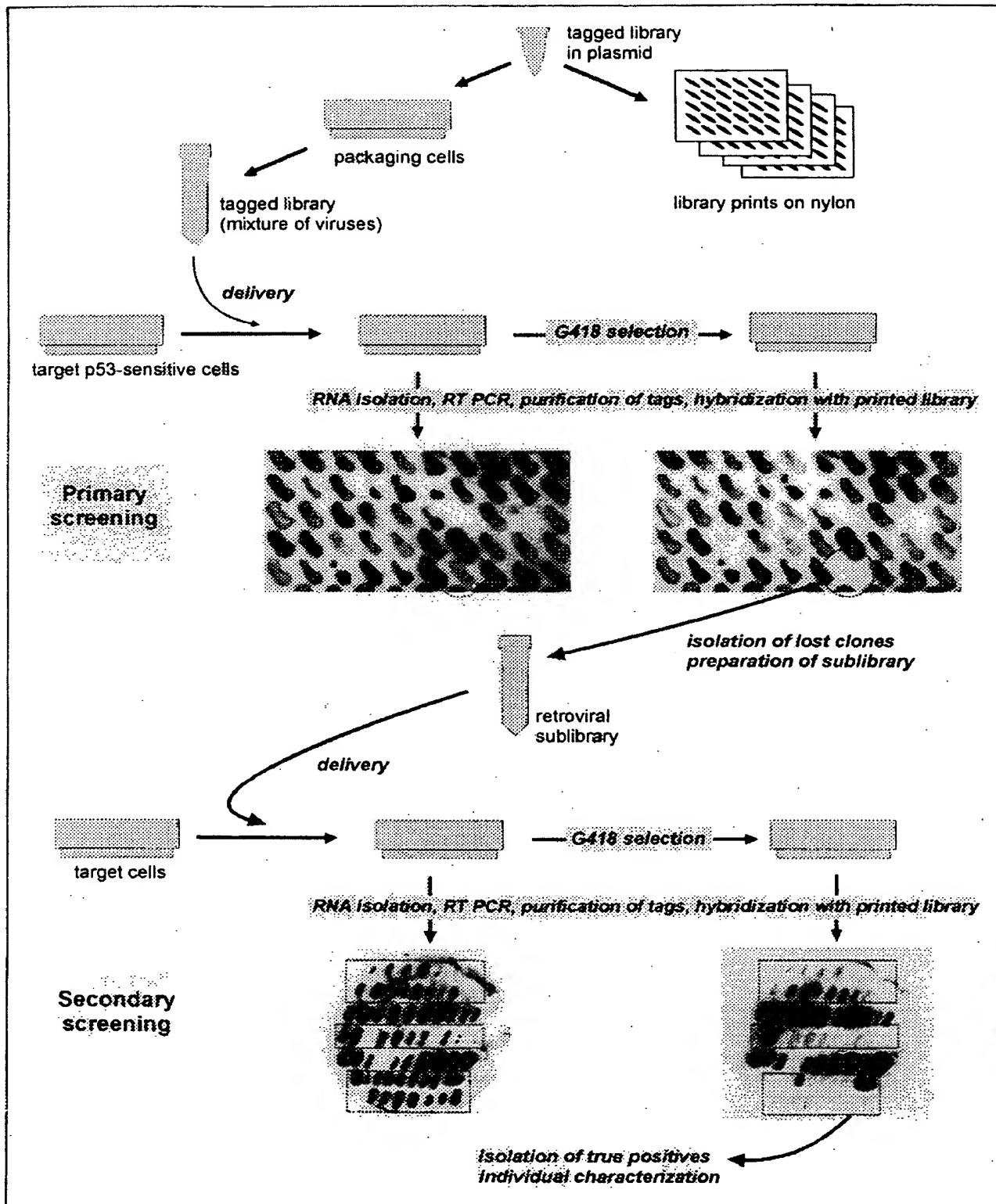
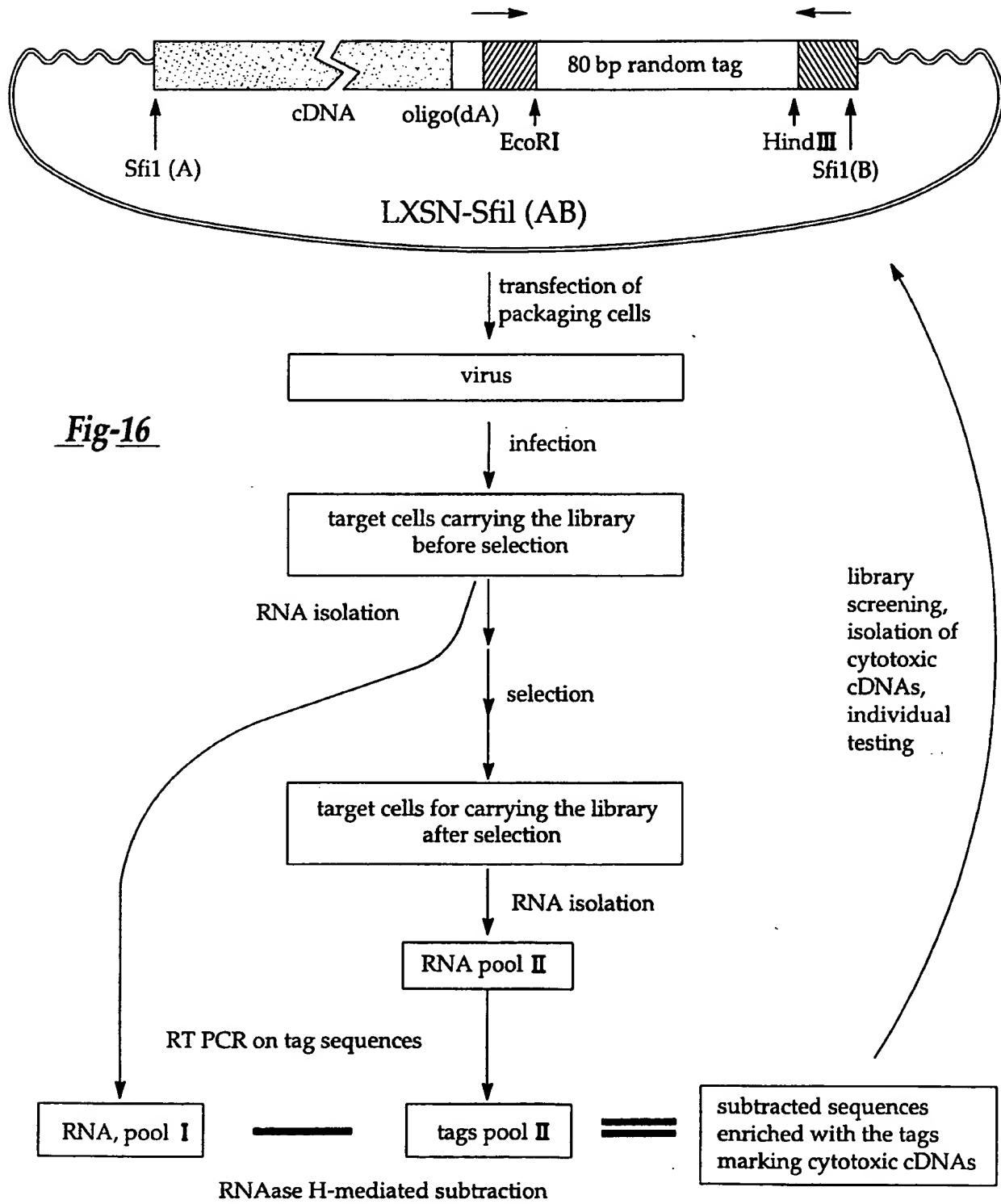
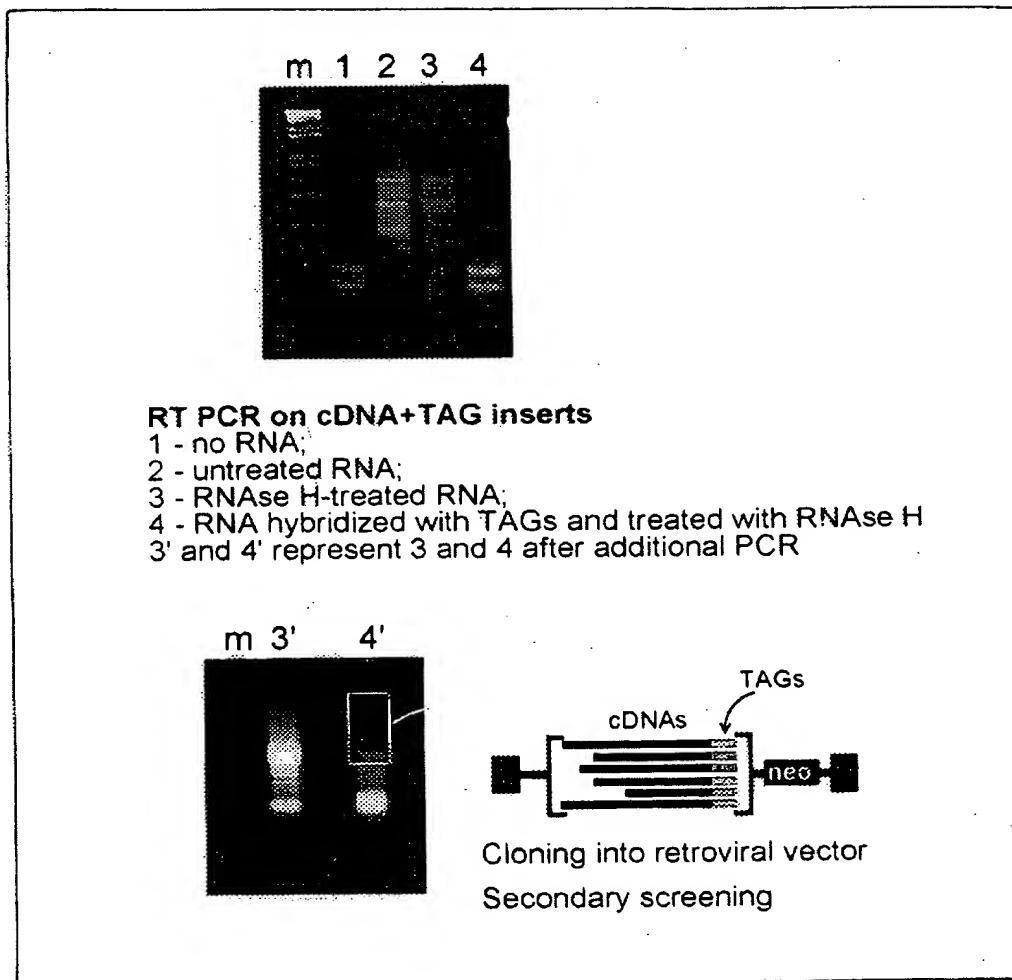


Fig-15

17/23



18 / 23

Fig-17

19/23

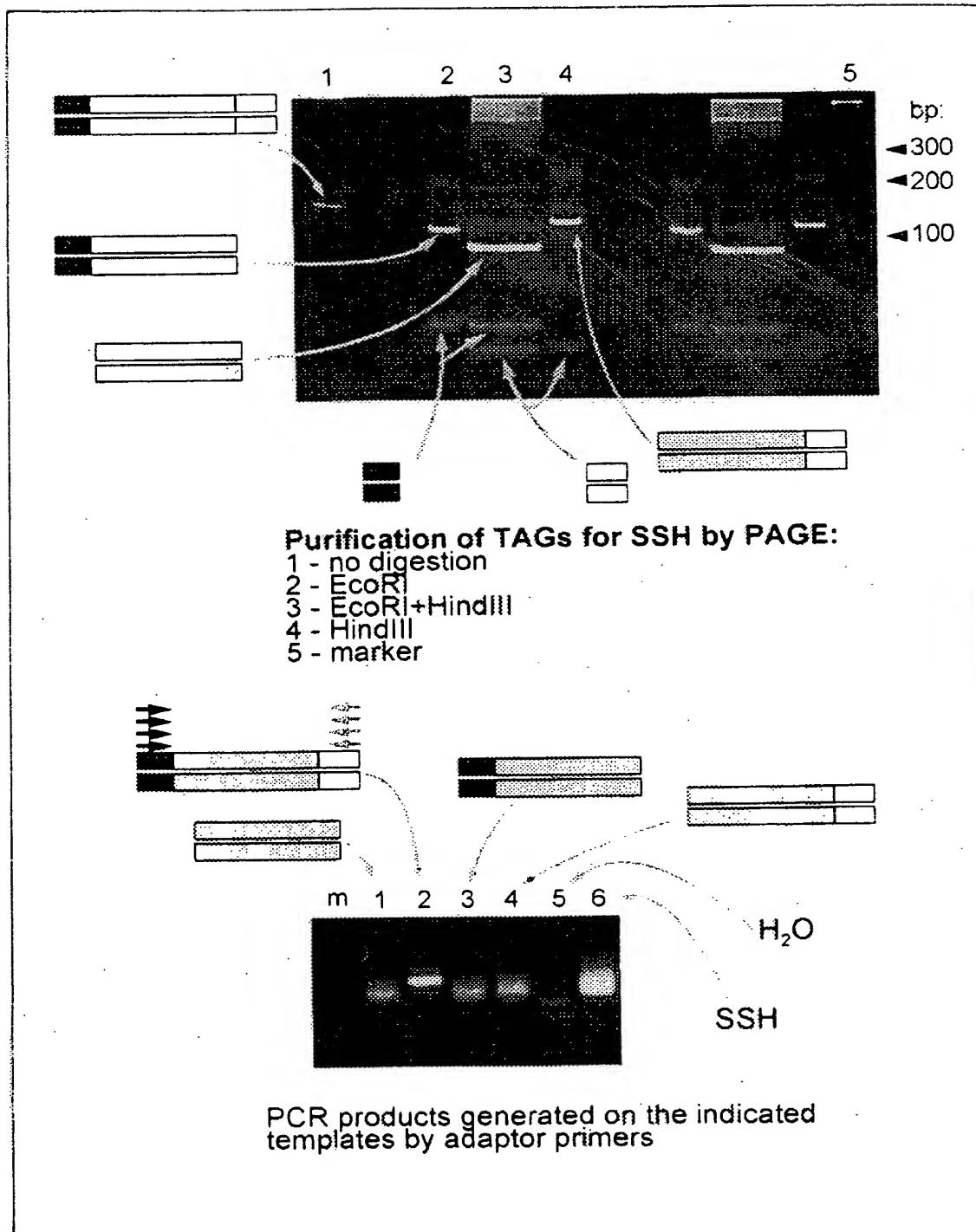
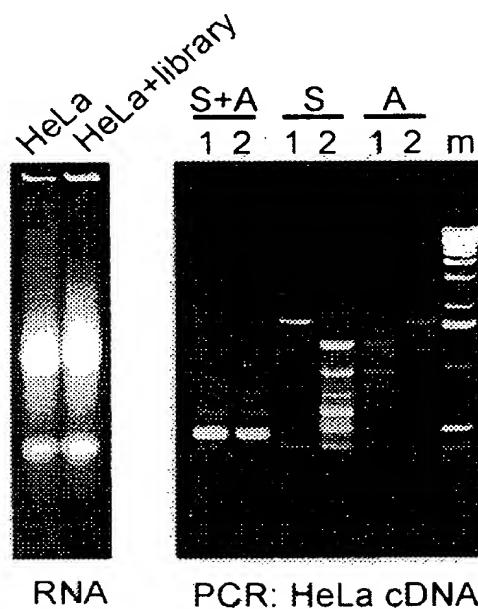
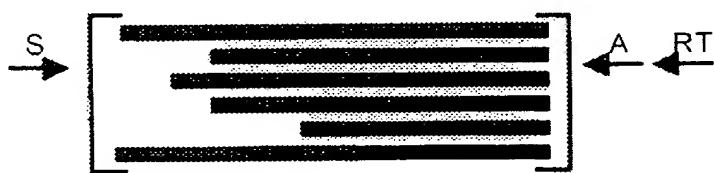
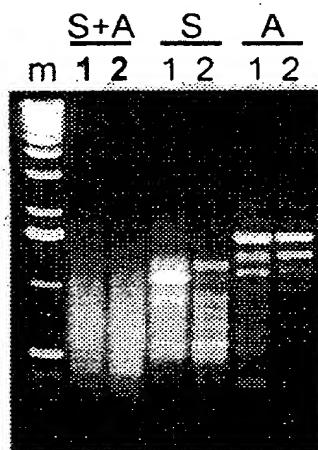


Fig-18

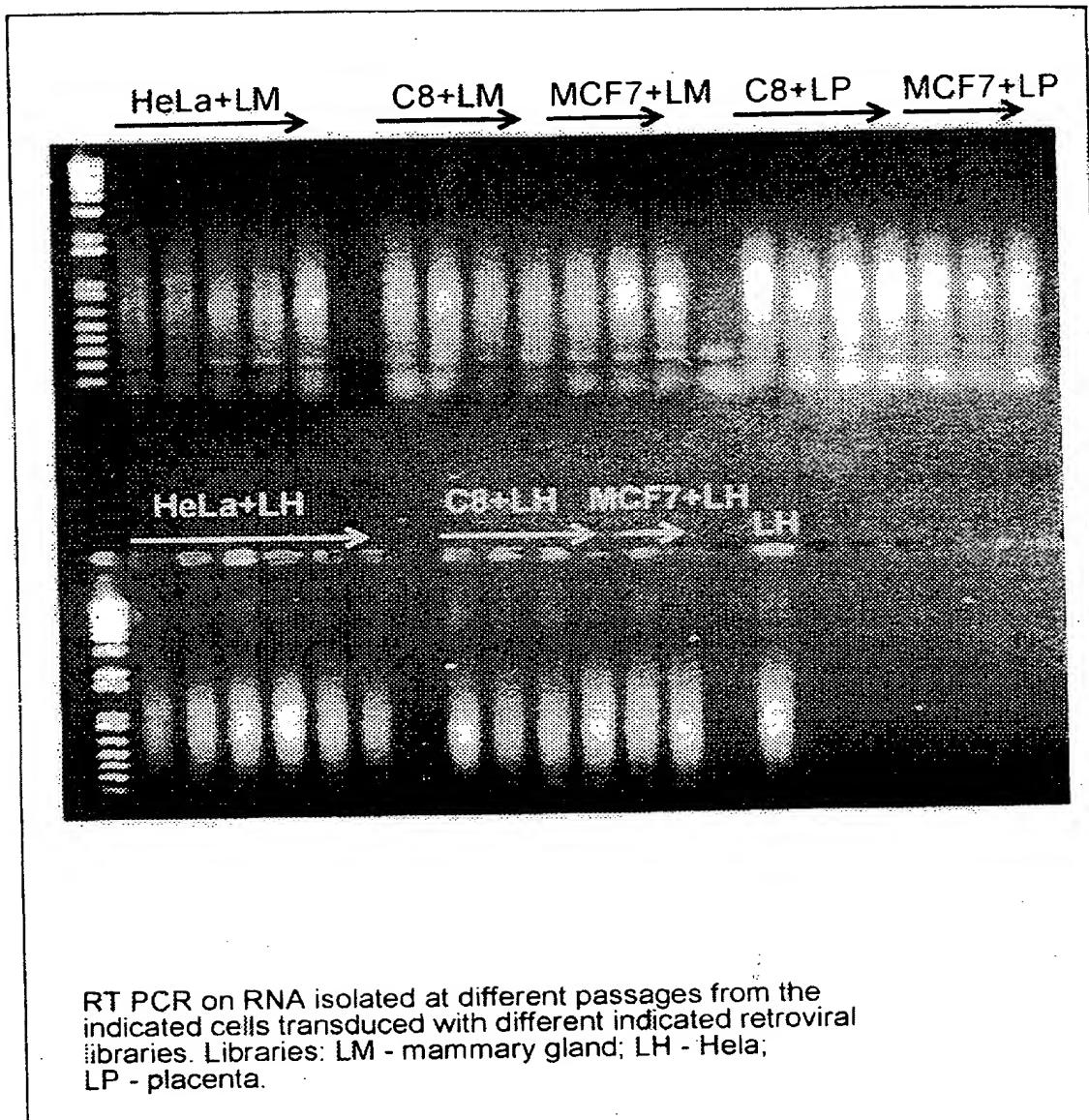
20/23

Fig-19

PCR: (HeLa+library) cDNA

- 1 - cDNA synthesized using vector-specific antisense primer
- 2 - cDNA synthesized using random primer
- S - sense-oriented primer
- A - antisense primer
- RT - vector-specific primer used for reverse transcription

21/23

Fig-20

22/23

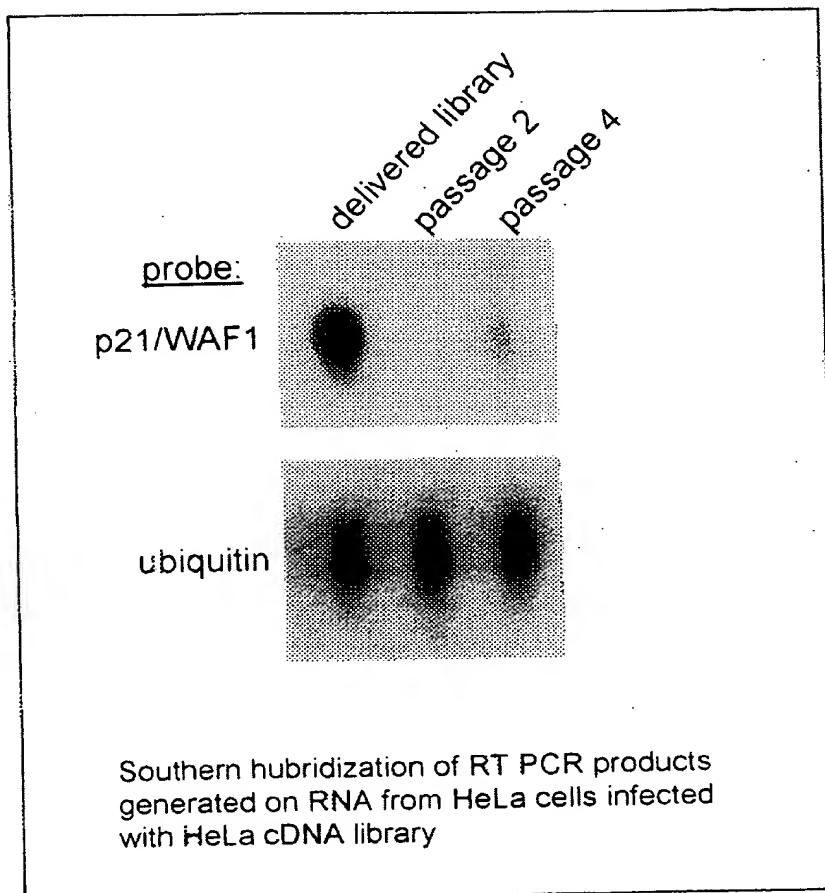
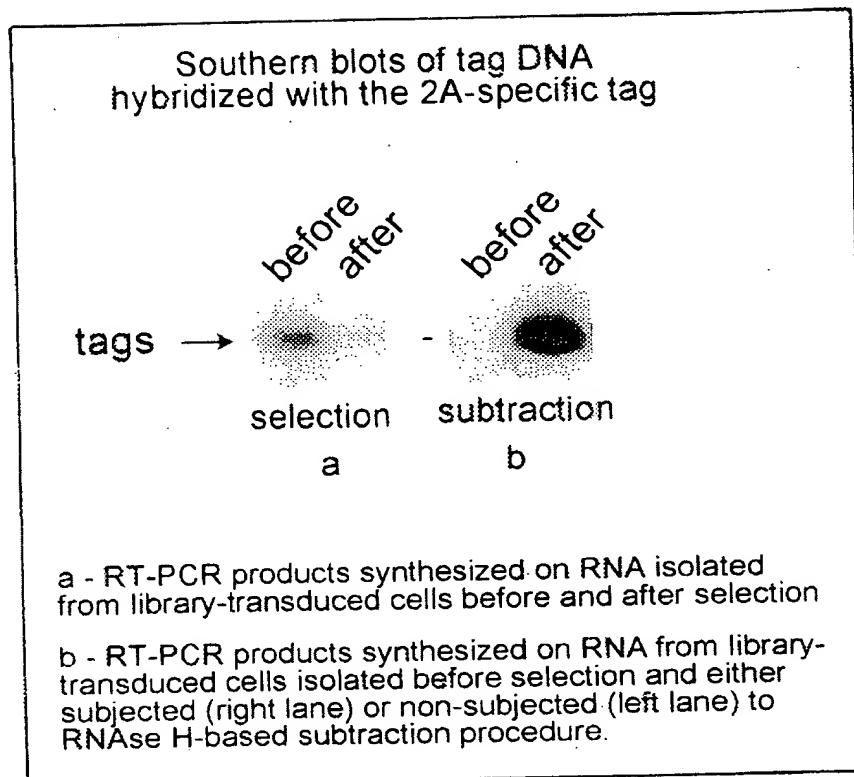


Fig-21

23/23

Fig-22

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US99/05944

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :C12N 5/00, 15/13; G01N 33/53  
US CL : 435/5, 7.1, 6

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/5, 7.1, 6

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, CAS, BIOSIS

Search terms: recombinant library, tag, cDNA-RNA hybrid

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5,580,717 A (DOWER et al) 03 December 1996, col. 5, line 50 up to col.6, line 52.	1-23
Y	US 5,427,908 A (DOWER et al) 27 June 1995, col. 5, line 65 to col. 6, line 68.	1-23
Y	DIATCHENKO, L. et al. Suppression subtractive hybridization: A method for generating differentially regulated or tissue-specific cDNA probes and libraries. Proc. Natl. Acad. Sci. USA. June 1996, Vol. 93, pages 6025-6030, entire document.	1-12, 14-19, 23

Further documents are listed in the continuation of Box C.

See patent family annex.

* Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"B" earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means		
"P" document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search

04 JUNE 1999

Date of mailing of the international search report

28 JUN 1999

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231  
Facsimile No. (703) 305-3230

Authorized officer  
*D. Lawrence T.*  
T. WESSENDORF  
Telephone No. (703) 308-0196

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US99/05944

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CRAMPTON, J. et al. The isolation of cloned cDNA sequences which are differentially expressed in human lymphocytes and fibroblasts. Nucleic Acid Research. 1980, Vol. 8, No. 5, pages 6007-6017, entire document.	20-22